American Community Survey 5-Year Summary File Users Guide

Supplement for DataFerrett

Contents

ACS 5-Year Summary File User's Guide Supplement for DataFerrett

December 2011

Introduction

The American Community Survey (ACS) is an ongoing survey that provides communities with the information needed to plan investments and services. By collecting demographic data, state and local leaders can better determine how more than $400 billion dollars of federal funding should be used. The information gathered helps them make decisions on infrastructure and services such as highways, schools, emergency services, etc.

This supplement provides basic information for accessing the ACS 5-year summary file data through DataFerrett, making a table, creating a map and for creating a user defined extract. DataFerrett allows users to download, view and manipulate a wide variety of data sets. The ACS 5-year summary file consists of aggregated data meaning that it has already been summarized and added up. Because of this, aggregated data can't be modified or recoded in DataFerrett. Currently, the ACS 5-Year Summary File data for both 2005-2009 and 2006-2010 are available.

For more information on step-by-step functionality in accessing and creating extracts of the ACS 5-year summary file data, see the Hands-On tutorial under the "Tutorials" link on the DataFerrett home page.

NOTE: For additional information, check out "Important Features and Limitations of the Current Release of the ACS 2006-2010 Summary File".

Accessing DataFerrett for the American Community Survey 2006-2010 5-Year Summary File

How to access DataFerrett

1. To access the ACS, we will be using the production version of DataFerrett. When you arrive at the DataFerrett icon in Figure 1, there is also a link to the New Features Update.

Figure 1: Click to enter DataFerrett

2. After clicking on the DataFerrett icon, you will arrive at the following screen (Figure 2).

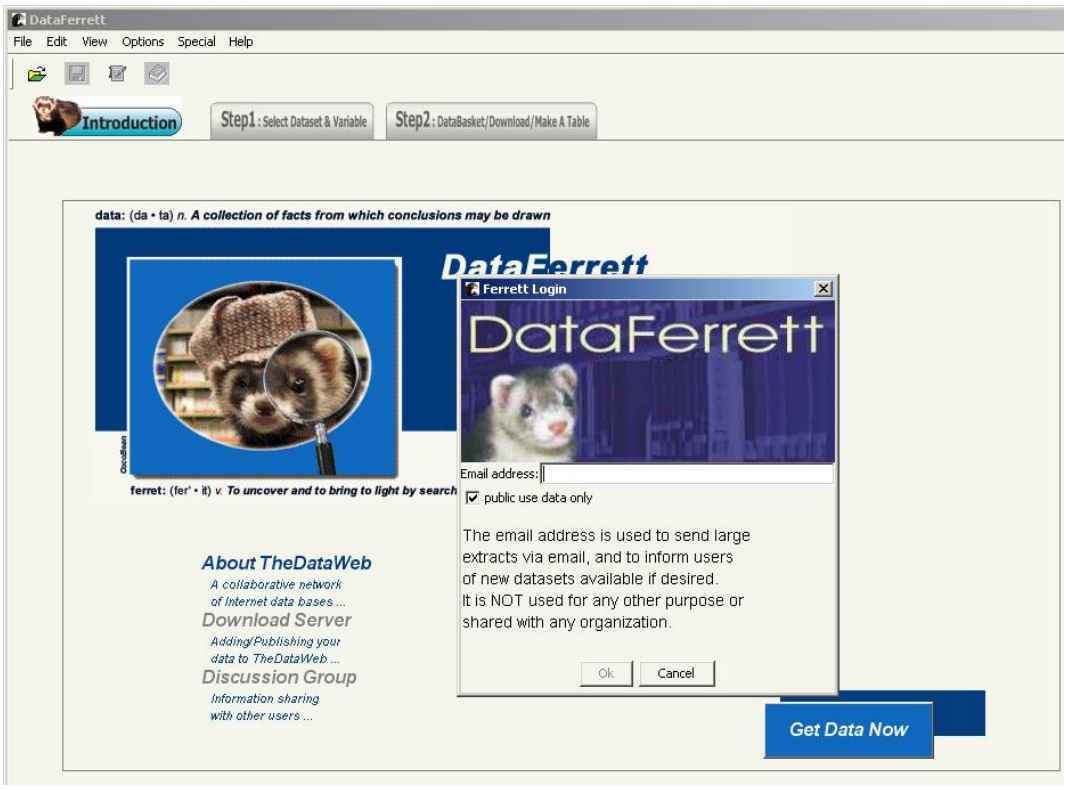

Figure 2: DataFerrett Login Screen

3. In order to log in to DataFerrett, you must provide a valid email address. After logging in and clicking on "Get Data Now", you will arrive at the Step 1 screen (shown in Figure 3), where you can select variables of interest.
4. On the far left side of the Step 1 screen is a listing of folders, also known as the Data Tree. When you locate the American Community Survey folder, click on the "plus" sign next to it.

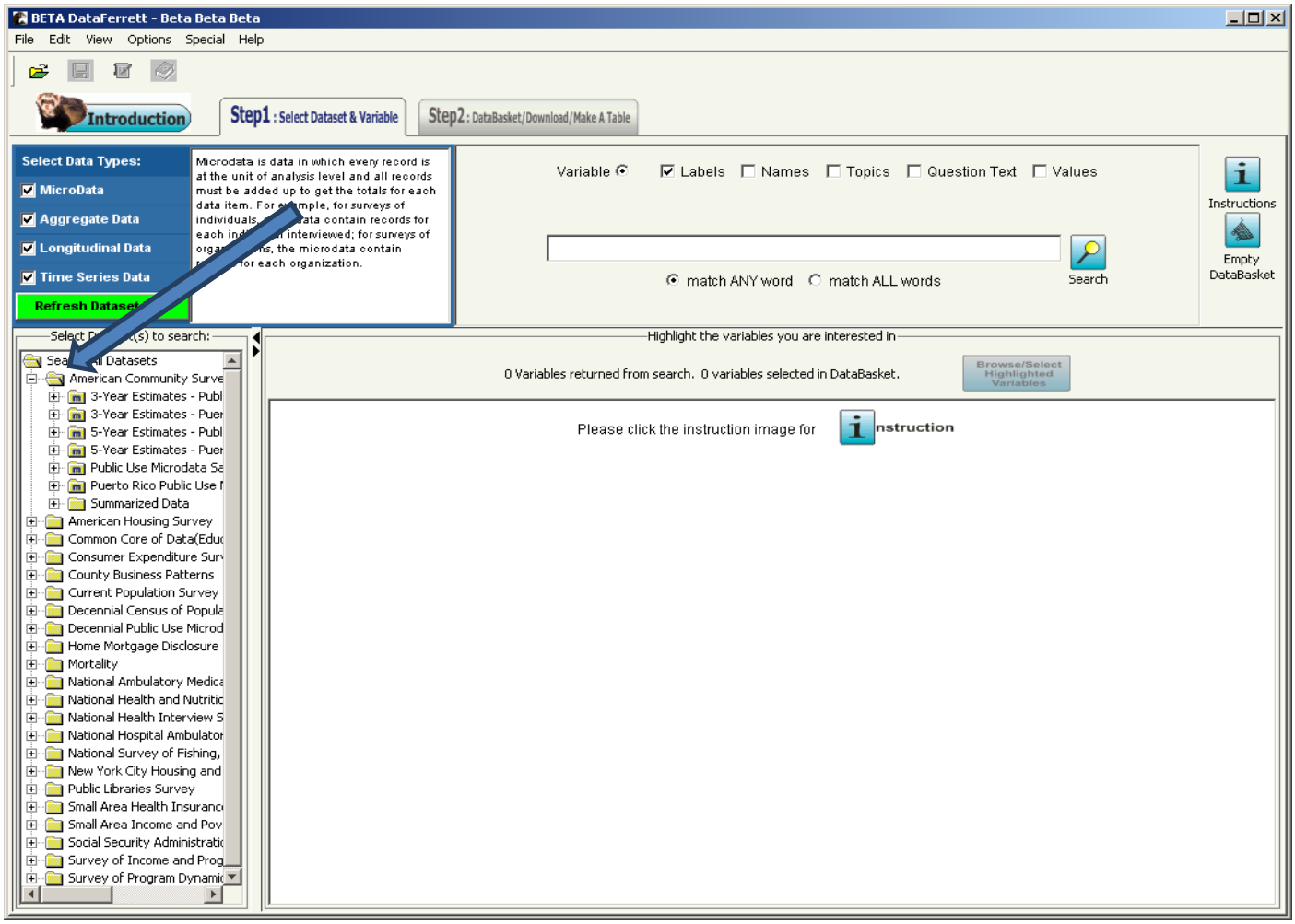

Figure 3: American Community Survey Folder

5. Continue clicking on the "plus" sign next to Summarized Data in the DataTree until you have opened the 5-year Summary File data as shown in Figure 4:

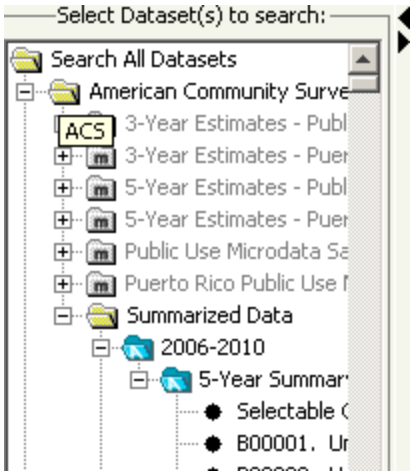

Figure 4: After opening all folders under Summarized Data

6. After choosing your data set and clicking on "Browse/Select Highlighted Variables", a new window will pop up displaying the ACS Summary 5-Year Notice. The 5-Year Notice is a brief message containing the link to Important Features and Limitations of the Current Release of the ACS 2006-2010 Summary File.

Example 1 – Downloading all cells in table B11001 for Alameda County, California

After clicking the plus sign next to the 5-Year Summary File folder, you will see a listing of over 900 data files. In this example, we will be looking at Household Type (Including Living Alone).

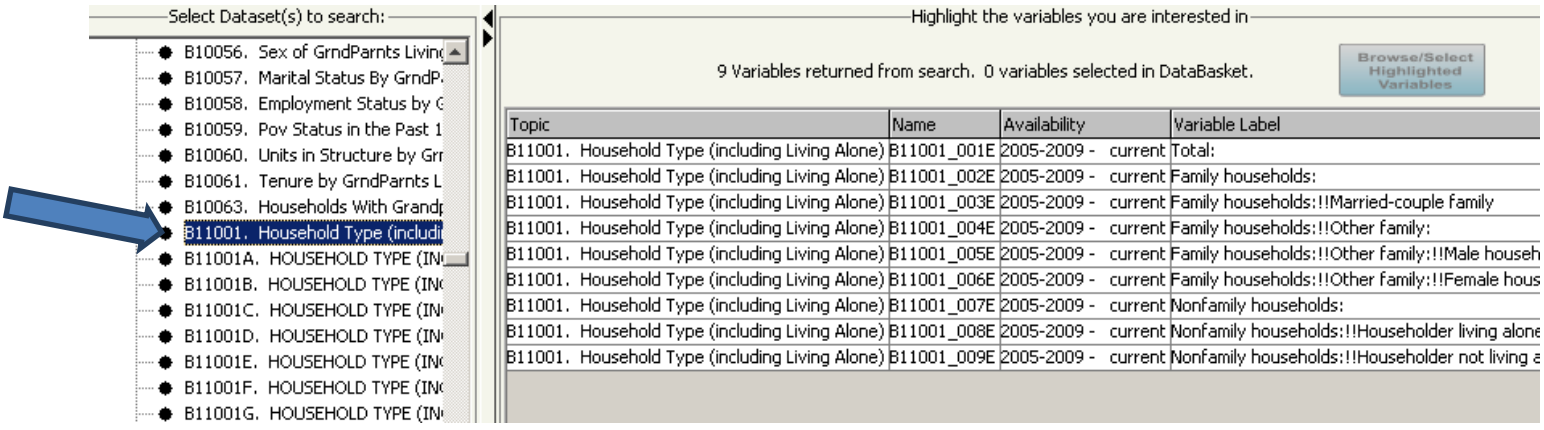

Figure 5: Household Type Variables

1. For this example, we will select all variables under Household Type by highlighting the first variable in the list and then clicking on the last variable in the list while holding down the "Shift" key.

2. When all variables are highlighted, click on “Browse/Select Highlighted Variables.”

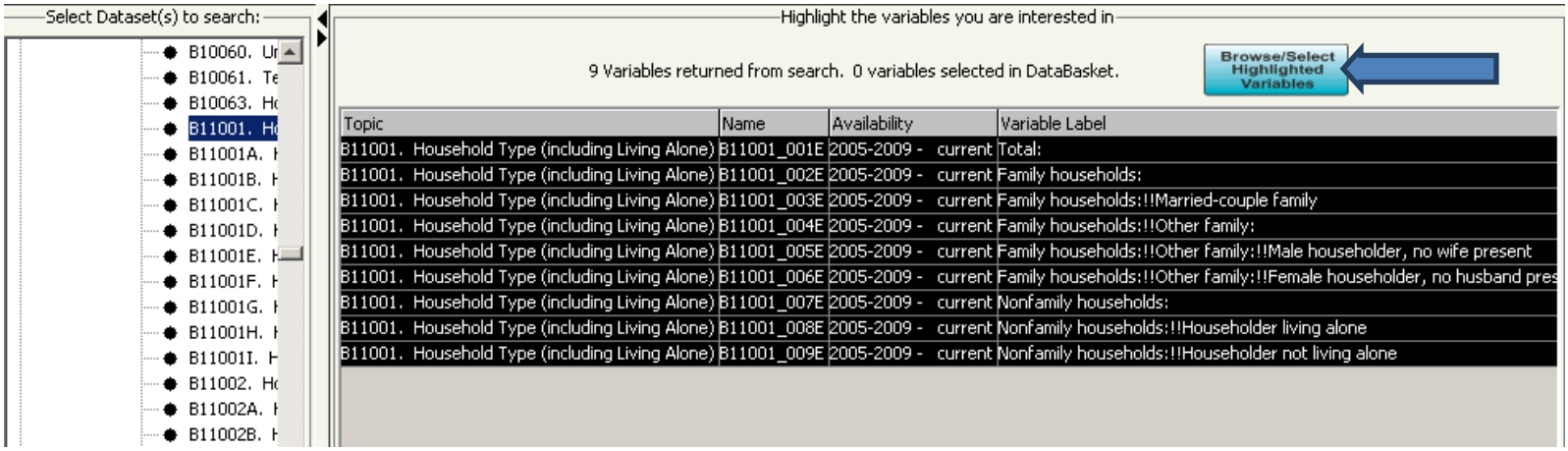

Figure 6: Highlight All Variables Under “Household Type”

3. Once the following window appears (Figure 7), click on “Select ALL Variables” and then “OK”.

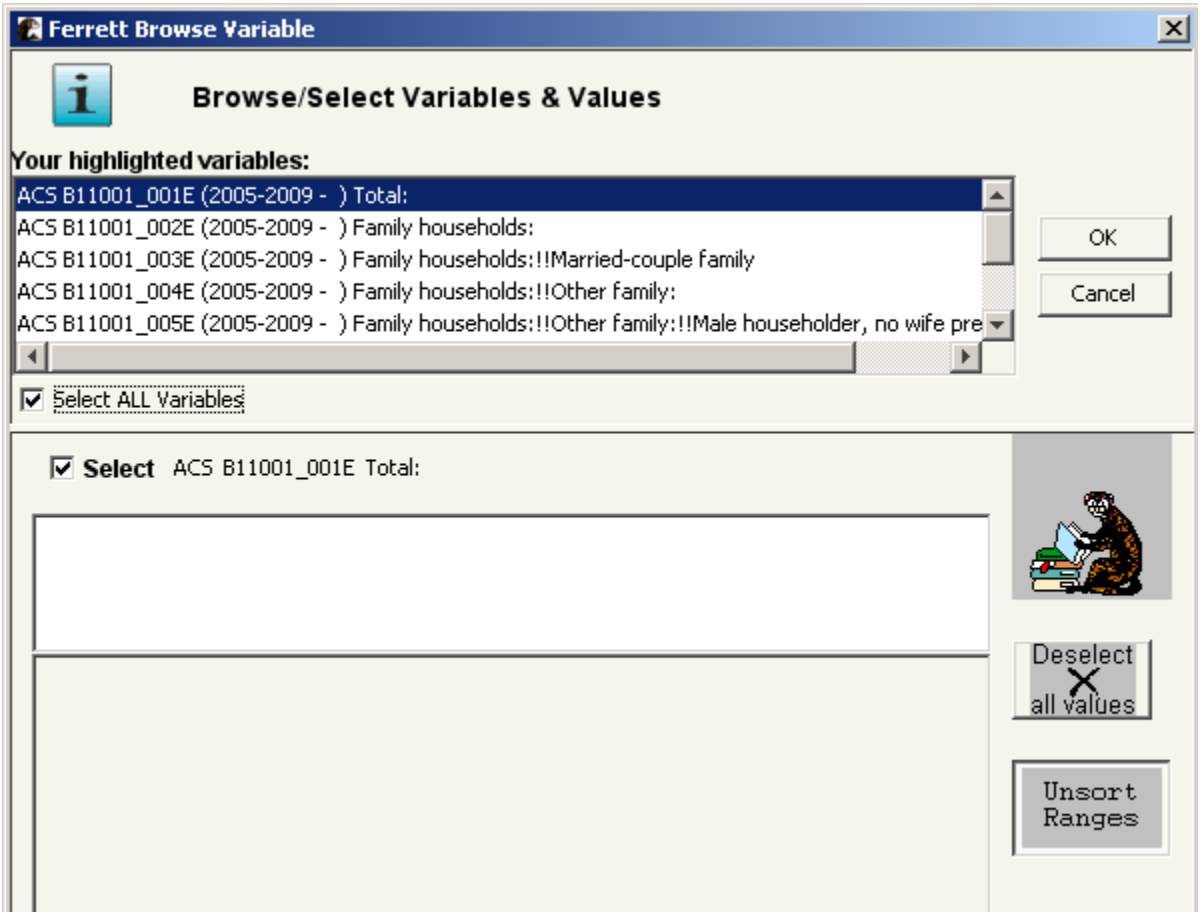

Figure 7: Select All Variables

4. After selecting all the variables (i.e., all cells in the table), you will see the following window:

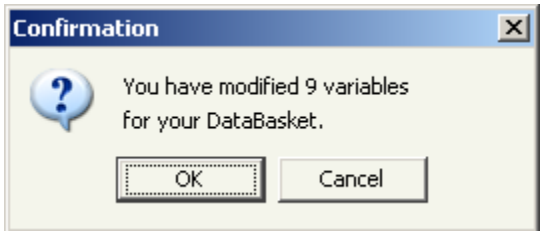

Figure 8: DataBasket Confirmation

5. After clicking "OK" the following "Ferrett Warning" window will appear:

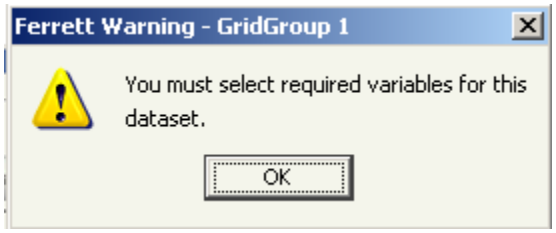

Figure 9: Ferrett Warning

Clicking OK on this pop-up window takes you to the screen displaying required and regular variables in Figure 10 at the top of the next page. This screen is where you start selecting the geographic area(s) you want.

6. User must select geographic areas

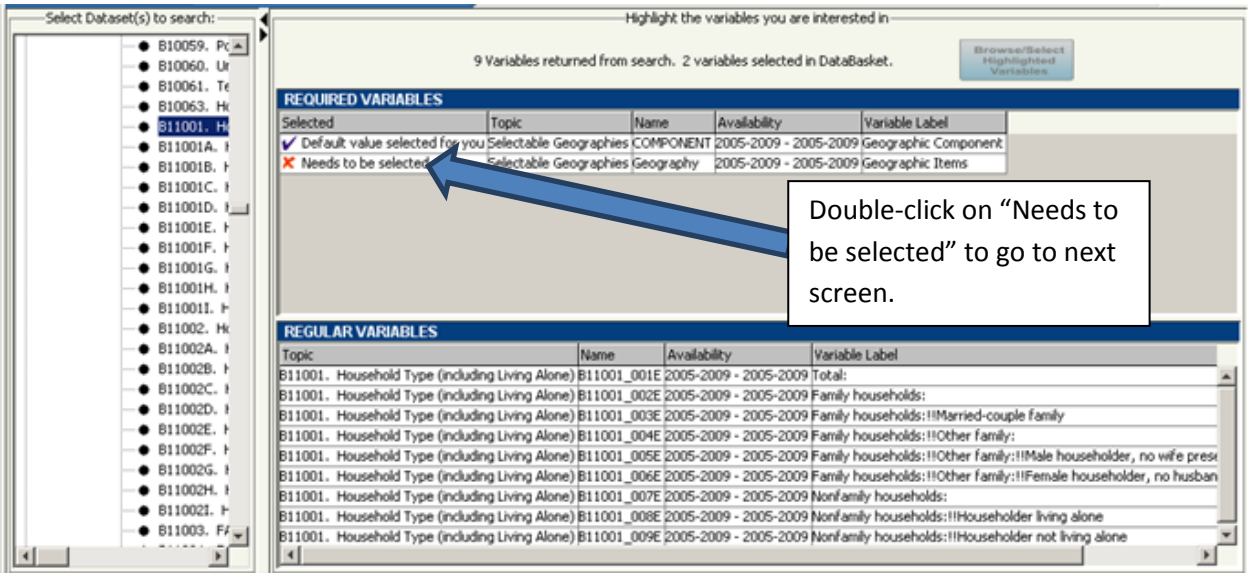

Figure 10: Screen showing geography that needs to be selected

7. Double clicking on "Needs to be selected" brings up the Ferrett Geography Codebook where the desired geographic areas can be selected. From here, follow the steps depicted in Figures 12 through 15.

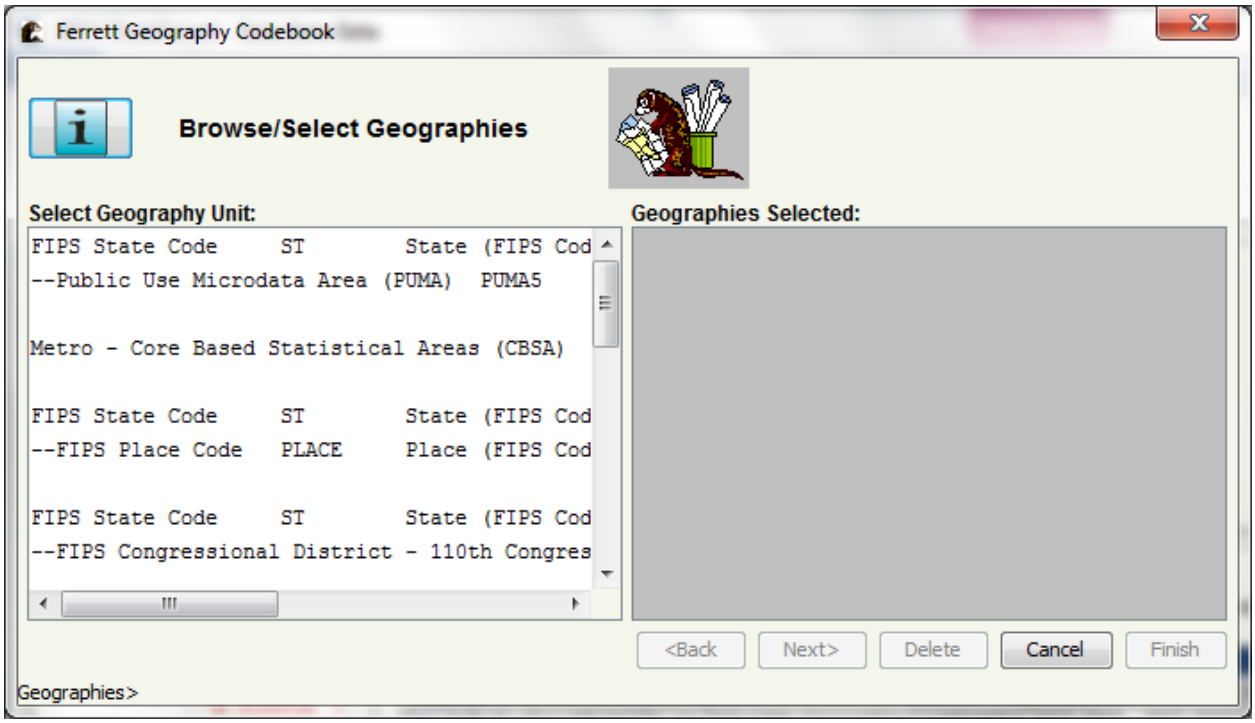

Figure 11: Ferrett Geography Codebook Screen – scroll down to "FIPS County Code" as shown in Figure 12.

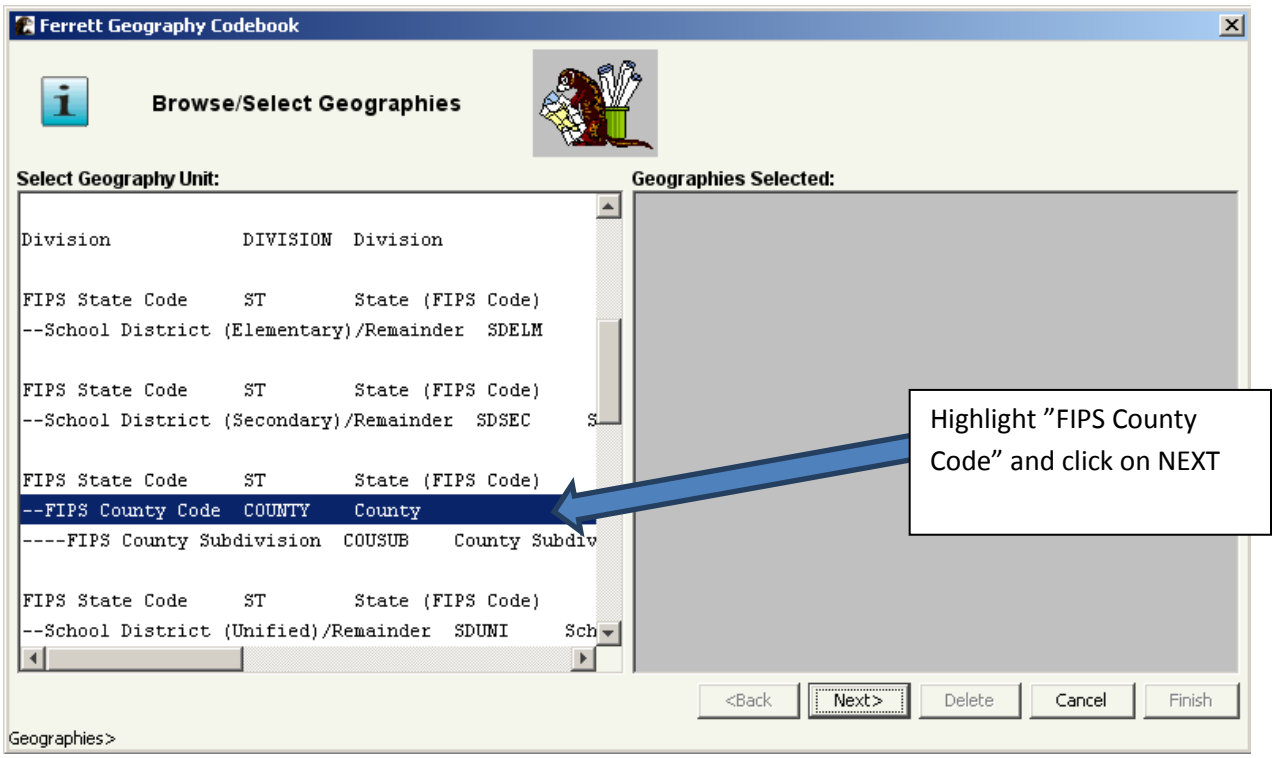

Figure 12: Select Geography FIPS County Code

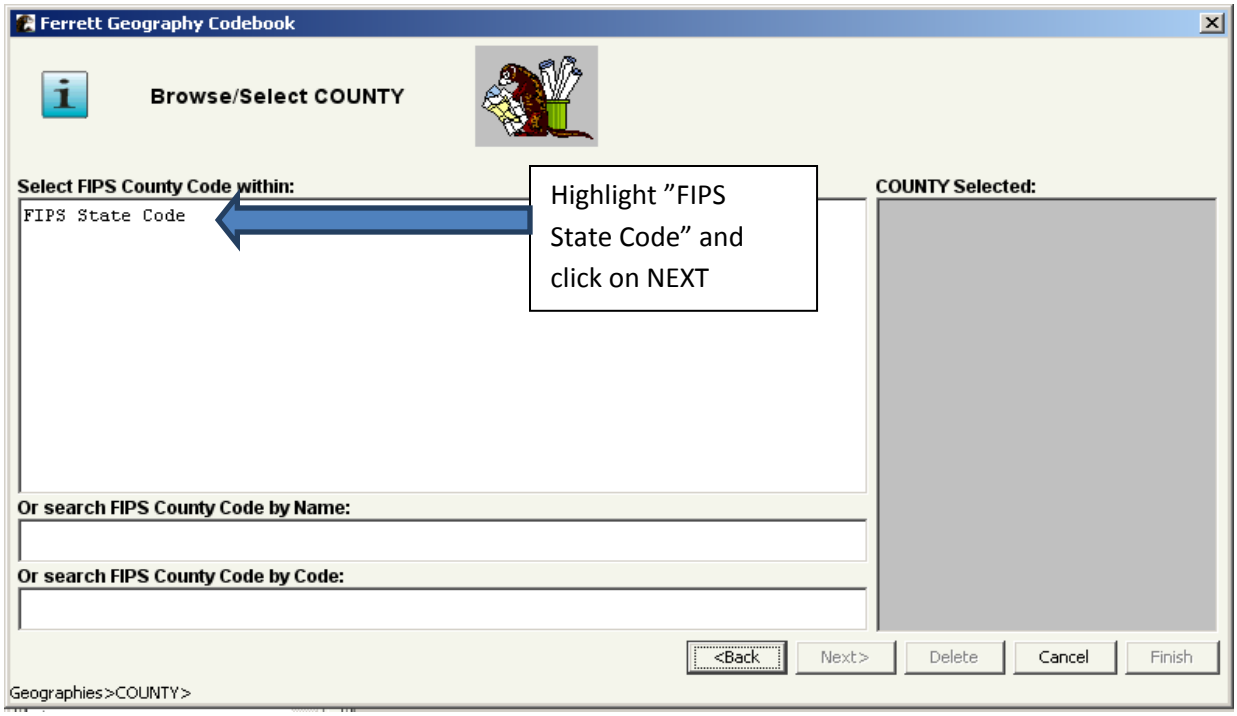

Figure 13: FIPS State Code

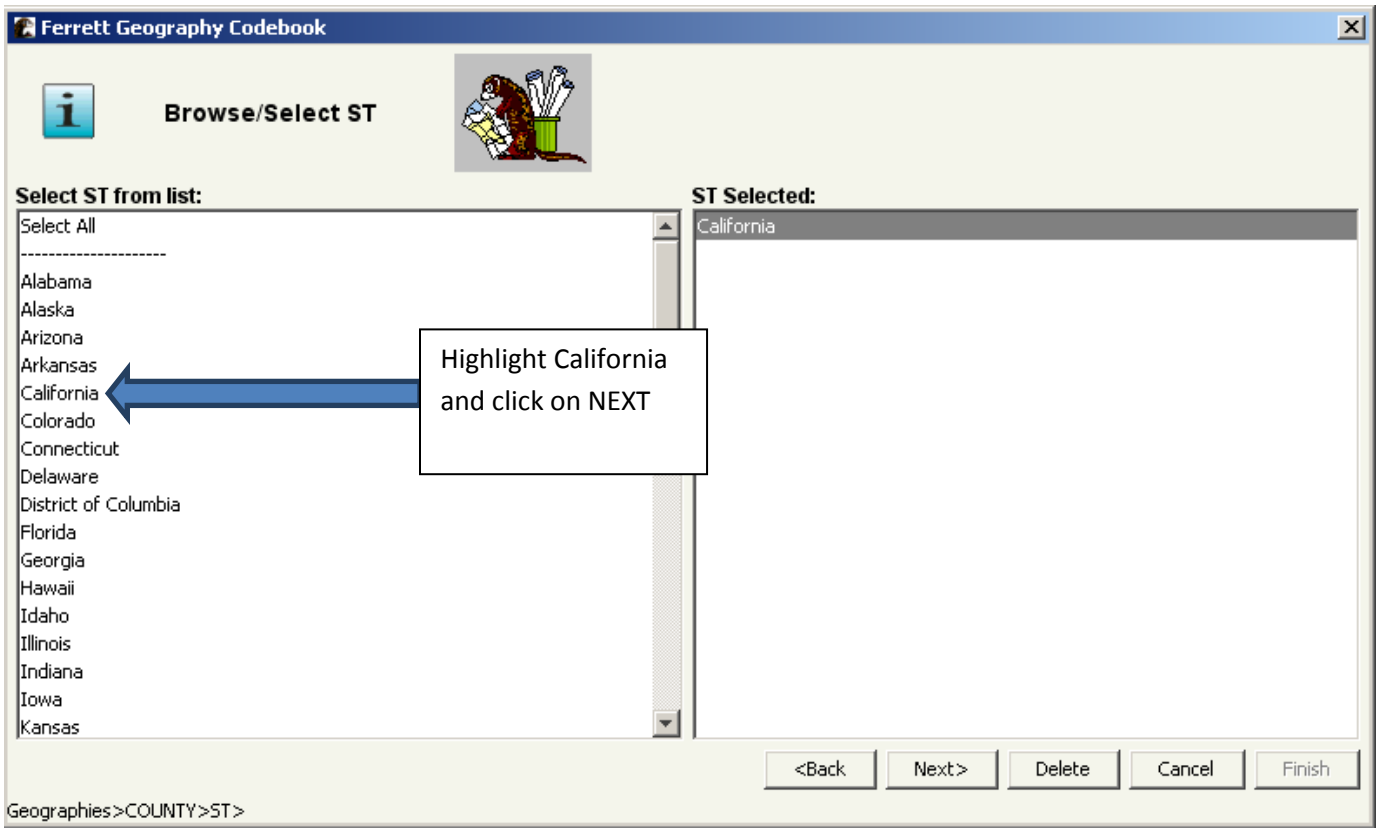

Figure 14: Select State From List

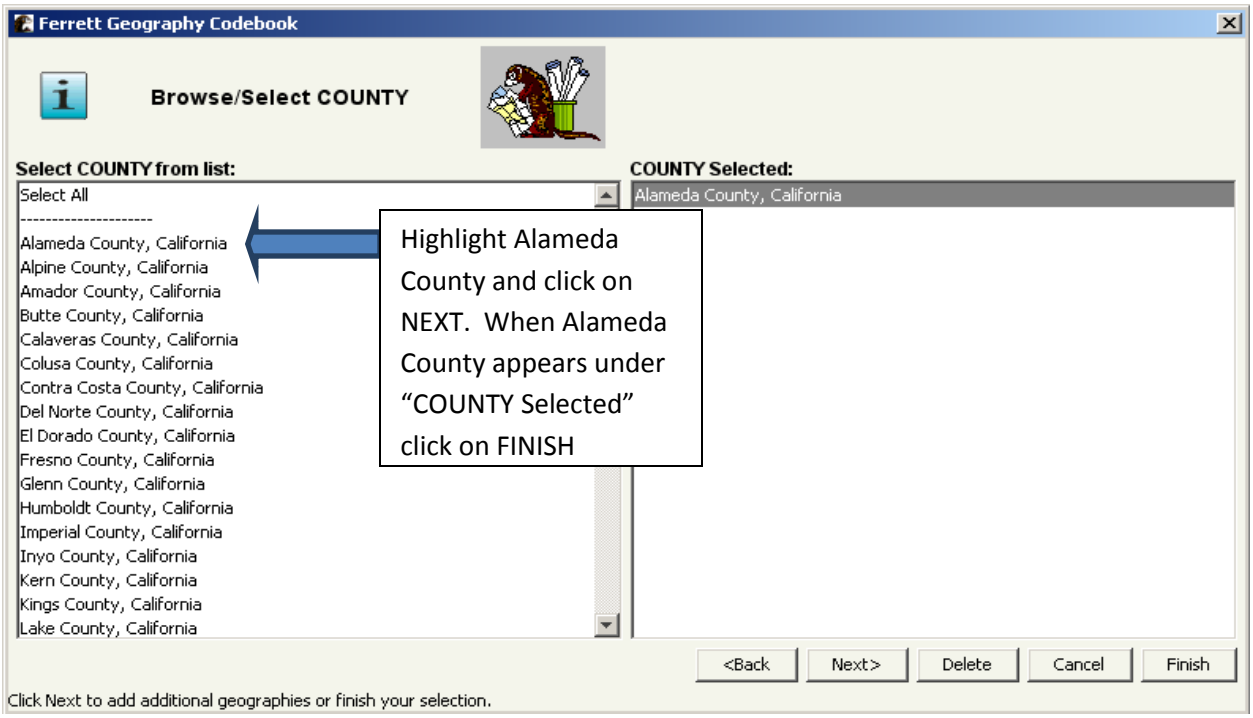

Figure 15: Select "Alameda County, California" then "Finish"

7. Now that you have all the needed variables, you can move on to "Step 2: DataBasket/Download/Make a Table". You will notice two choices at the top of the screen: "Download" or "Make a Table". Click on the "Download" icon shown in Figure 16.

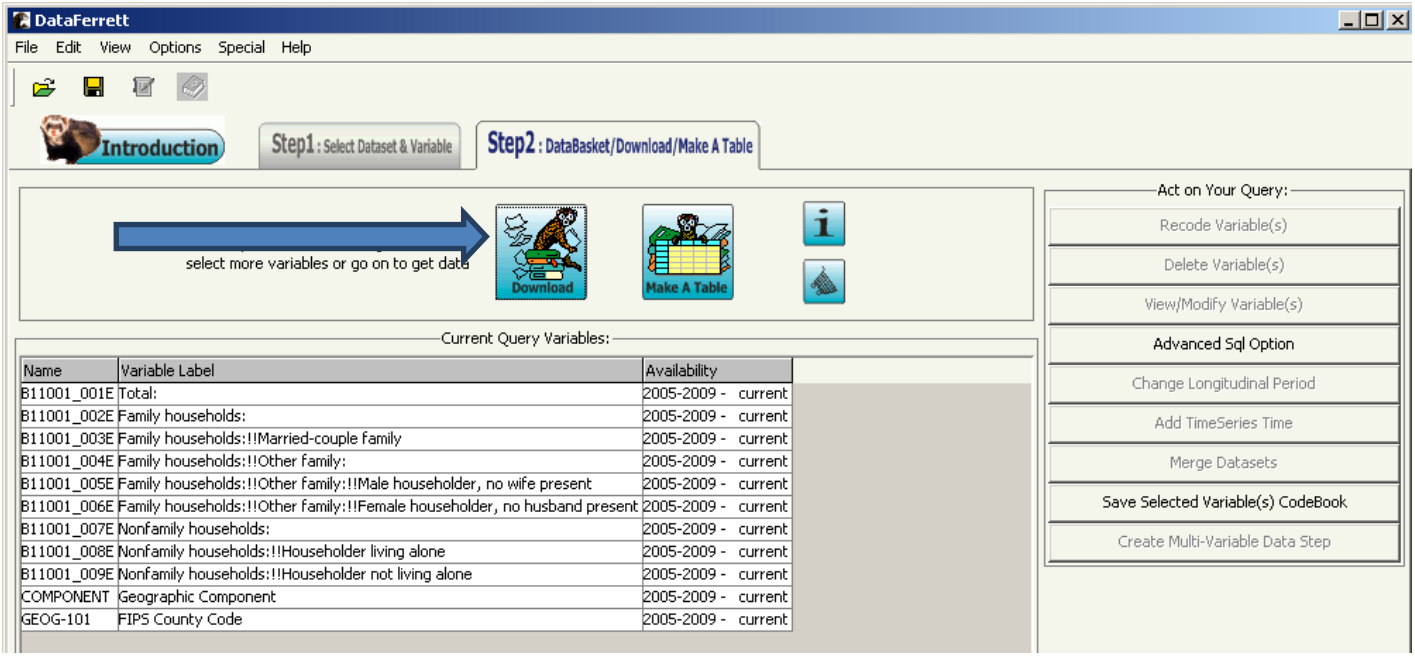

Figure 16: Click "Download" Icon

After clicking on "Download" the "Download Data" window appears (Figure 17). By choosing "Download" in Step 2, the "Download data in batch mode?" check box is automatically selected. Because "batch mode" is selected, once processing of the extract request has completed, a message will be sent to the email address you provided when you logged in.

Another feature (highlighted in Figure 17) allows you to include Margins of Error (MOE) in your selection, a feature new to the latest version of the ACS 5-year Summary File.

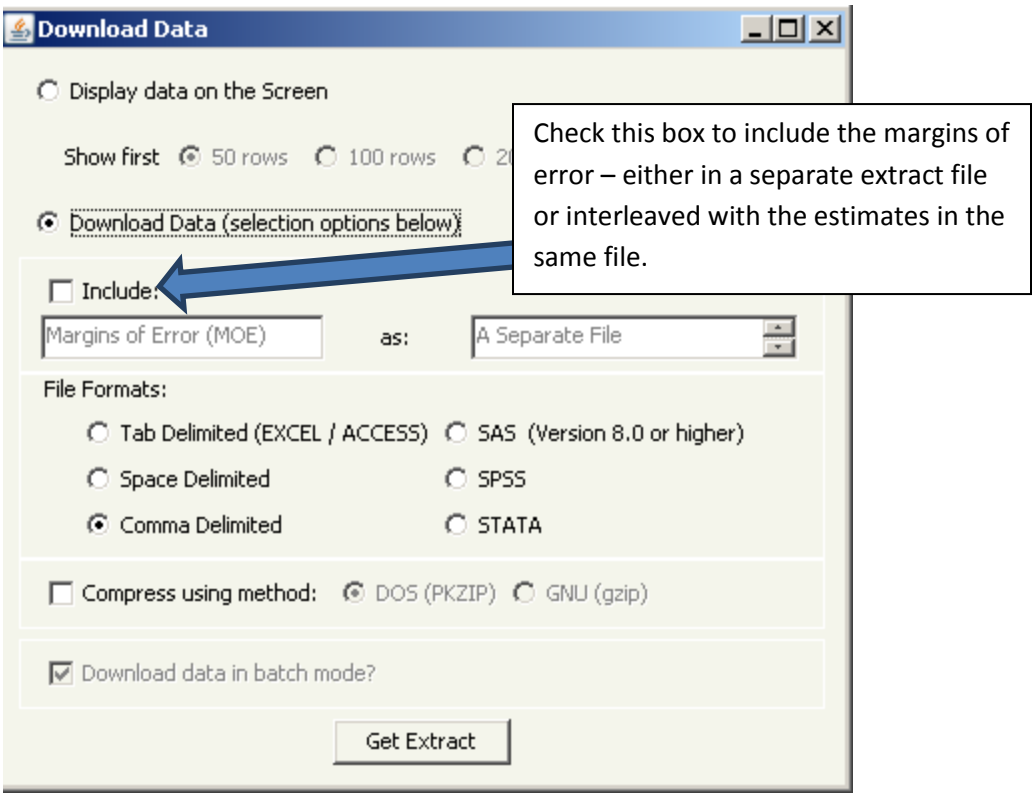

Figure 17: Download Data Screen with Margin of Error Feature

When attempting to download data in "batch mode", you may find that you have exceeded an extract size threshold. The following four images depict Ferrett Extract Warnings that can appear when you exceed these thresholds.

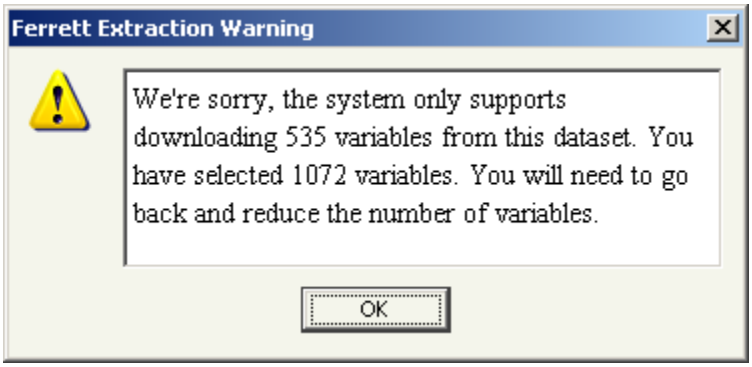

Figure 18 Warning about exceeding maximum number of variables

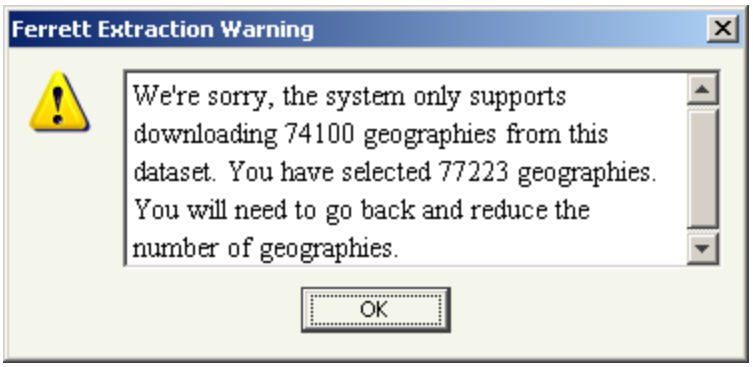

Figure 19: Ferrett Extraction Warning Limited to Specific Number of Geographies

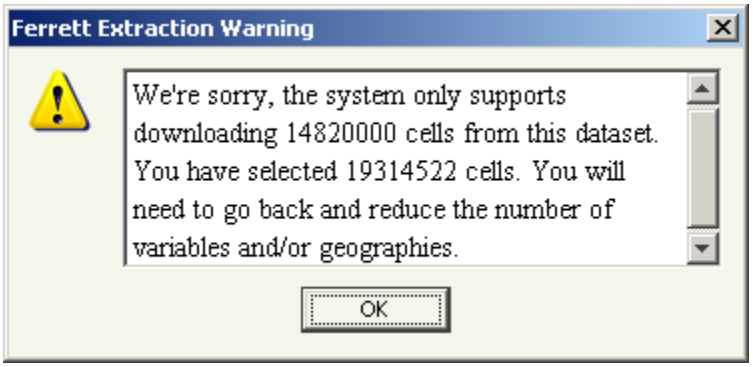

Figure 20: Ferrett Extraction Warning Limited to Number of Cells Selected

Figure 21: Ferrett Extraction Warning Prompting User to Select File Compression Method

If there are no issues with the extract size threshold, the message "*Please Wait… Processing your Request*" will appear after clicking on "Get Extract" which will then be followed by this pop-up:

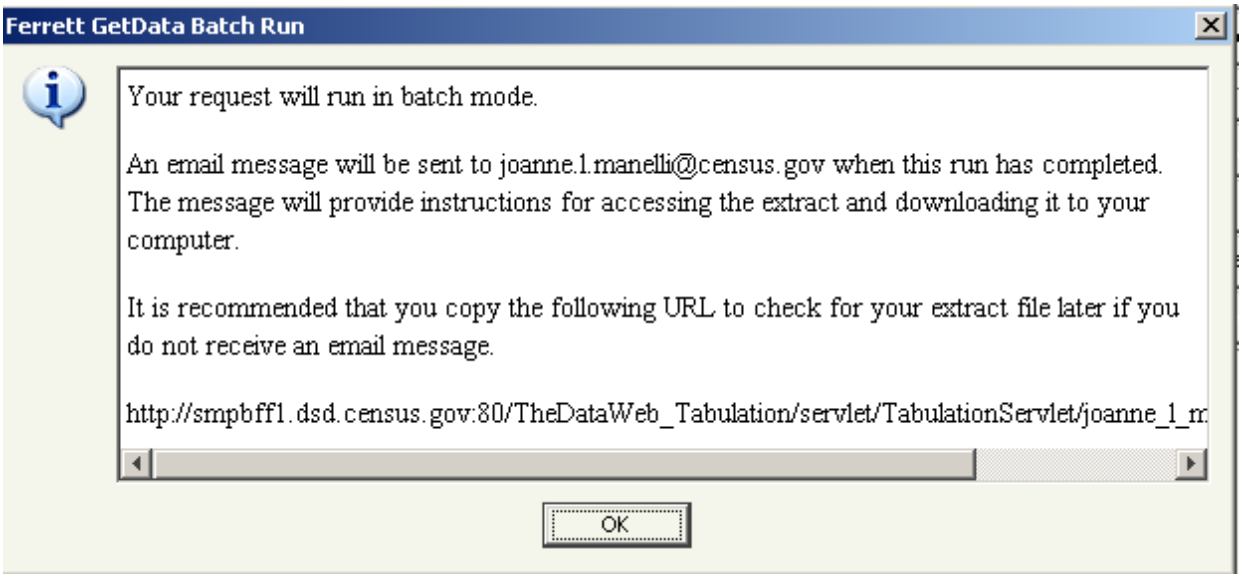

Figure 22: Ferrett "GetData Batch Run" Pop-up Message

After clicking "OK" check your email and you should find a message stating that your extract has been created and is ready to be downloaded to your computer. If you do not see this message within a few minutes, check your "Spam" or "Junk" email folder. Clicking on the URL in that message will take you to the extract page. Below (Figure 23) is an example of what the extract page will look like.

(NOTE: A sentence is included at the bottom of this screen reminding you that you should save the "Codebook" to your computer. This text contains descriptions and labels of the variables in your DataBasket.)

Extract Request for: joanne.l.manelli@census.gov

NOTE: Use the right mouse button on the file link(s) to "`Save Target As:/ Save Link As:`", which will allow you to download your file(s) to your computer.

Extracted: 1 record(s) for 2006-2010

- Comma Delimited ASCII File for 2006-2010 joanne_1_manelligvl8zkwo1.asc

Also note, the Record Layout File is not a substitute for a codebook. To obtain a codebook which contains variable descriptions and value labels, return to DataFerrett - Step 2 DataBasket... and click the button labeled "Save Selected Variable(s) Codebook".
Length of time: 0.1 minutes.
Created: Nov 29, 2011 1:36:12 PM

Contact Information:
E-mail: dsd.ferrett@census.gov

DataWeb Help Desk
Toll Free: 866-437-0171

Figure 23: Extract Request Received Through Email Message

Here is what the first few data records will look like in your downloaded file:

```
"GEOID," B11001_001E       B11001_002E       B11001_003E       B11001_004E       B11001_005E       B11001_006E       B11001_007E
05000US06001       520096    331446    241783    89663     26188     63475     188650    147345

B11001_009E       SUMLEVEL  ST        COUNTY    COMPONENT
41305     50        6         1         0
```

Figure 24: Data Record Download File

Below is an example of what the data records will look like after loading them into a spreadsheet application like Excel.

GEOID,	B11001_001E	B11001_002E	B11001_003E	B11001_004E	B11001_005E	B11001_006E	B11001_007E	B11001_008E
05000US06001	520096	331446	241783	89663	26188	63475	188650	147345
B11001_009E	SUMLEVEL	ST	COUNTY	COMPONENT				
41305	50	6	1	0				

Figure 25: Data Record in Excel File

A few things to note about the output file:

1. DataFerrett treats every value as numeric. Instead of county “001” in state “06”, you will see county 1 and state 6.
2. DataFerrett will not automatically output the label, or name, of the geographic area, i.e. “Alameda County, CA”. To add the label/name to your output, you must select the variable “NAME” which you will find in the list of geographic variables when you highlight “Selectable Geographies” on the left-hand side of the page in Step 1 (right above the first table, B00001).
3. The “Component” field refers to the geographic component which is always included in the output file. The default value of 0 (=”00”) is pre-selected in DataFerrett. Actual geographic components (e.g., “Rural” or “Urban”) must be selected by the user to override the default value (See pages 37-38 more detailed information.)

Example 2 – Downloading all cells in table B11001 for all Block Groups in Alameda County, CA

When searching geographic areas for variables, you may want to look at block groups within a census tract. Census tracts consist of relatively permanent statistical subdivisions within a county or statistically equivalent entity, delineated for data presentation purposes by a local group of census data users or geographical staff of a regional census center in accordance with Census Bureau guidelines. A block group is a statistical subdivision of a census tract consisting of all tabulation blocks beginning with the same digit. For example, in Census 2000, BG 3 represents all blocks numbered between 3000 and 3999.

1. To gather block group information, follow Steps 1 through 4 as described in Example 1. Once you reach the “Browse/Select Geographies” window, scroll down until you see “Census Block Group” (Figure 26) then click “Next”.

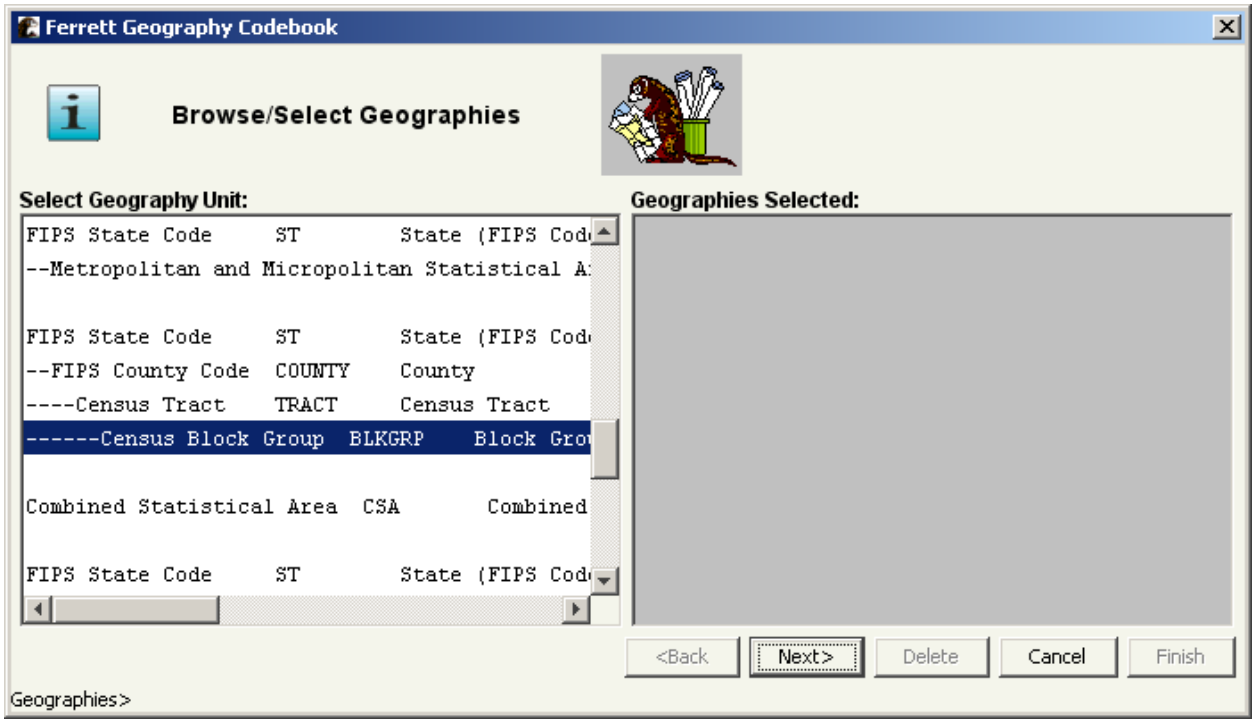

Figure 26: Select "Census Block Group" in "Browse/Select Geographies"

2. In this example, we will be selecting "FIPS County Code" (Figure 27).

Figure 27: Select "FIPS County Code"

3. After clicking “Next”, you will be brought to the following window (Figure 28). The only option listed will be “FIPS State Code”.

Figure 28: “FIPS State Code” within Block Group

4. Just like in Example 1, select California from the list of states.

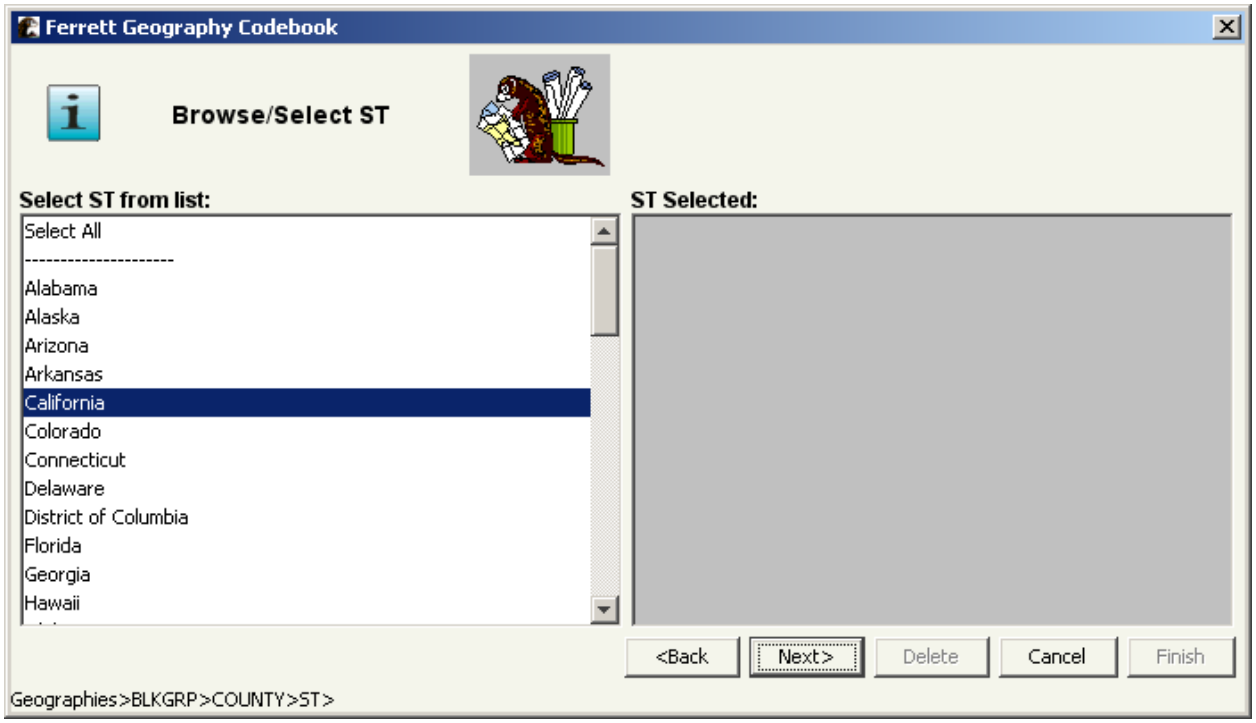

Figure 29: Select State from List

5. Continue selecting the state and county (See Figures 14 and 15) but this time, click on “Next” instead of “Finish”.

Figure 30: Select County from list

6. After clicking “Next” you will notice a window with just “FIPS State Code” listed but to the right “Alameda County, California” is listed as well.

Figure 31: Alameda County, California selected

7. Click “Next” to arrive at the complete listing of all block groups within the county (Figure 32). Highlight “Select All” then click “Next”.

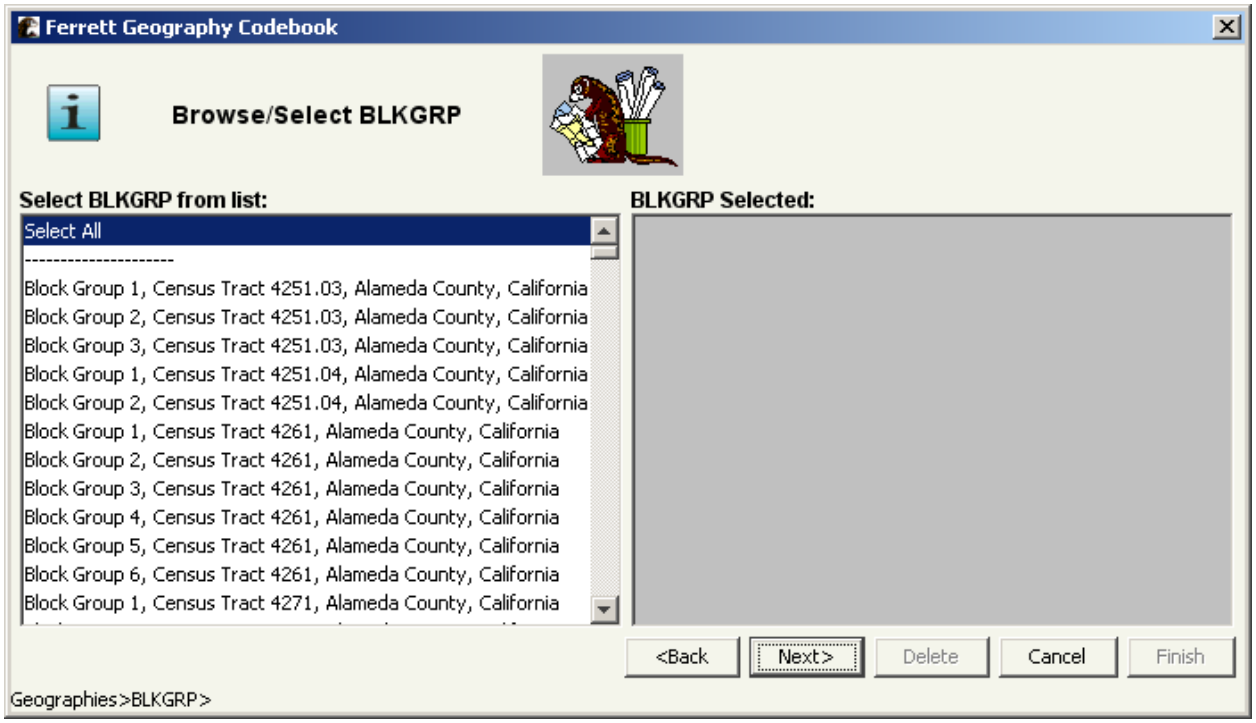

Figure 32: Block Groups within Alameda County, California

8. Click on "Finish" once the block groups appear in the "Selected" area on the right (Figure 33).

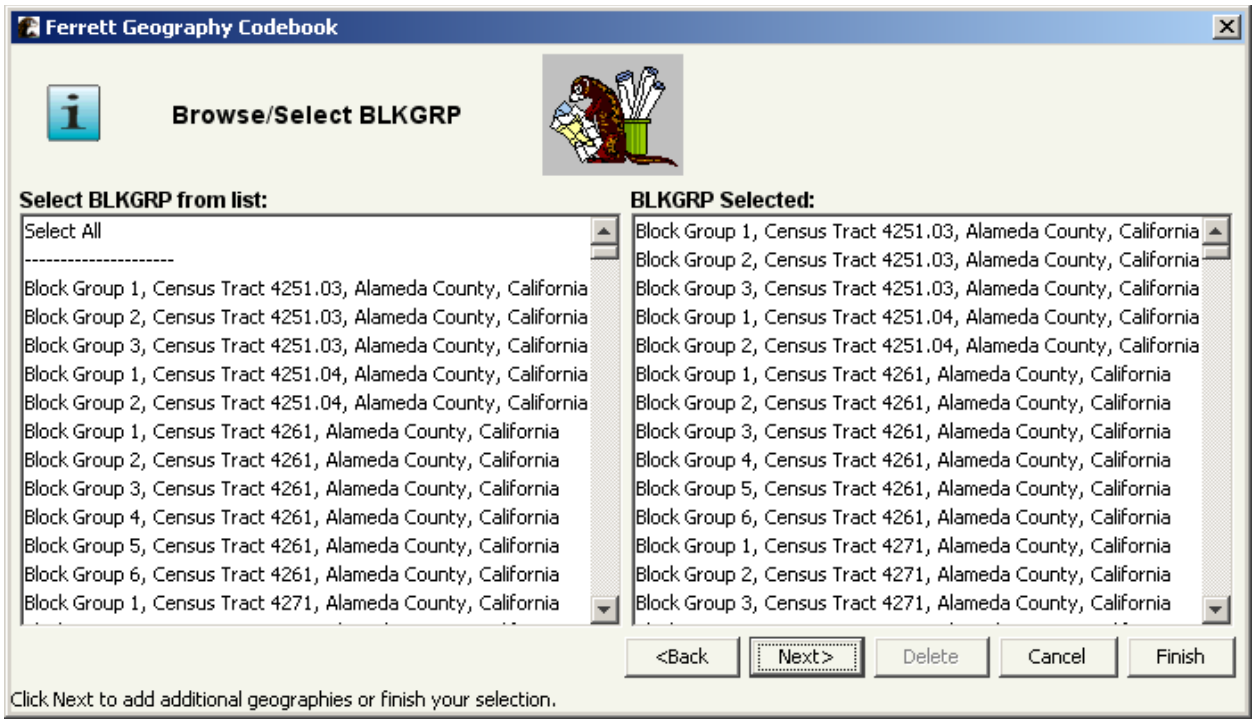

Figure 33: All Block Groups Selected

9. After you have finished selecting the block groups within Alameda County, California, follow the steps shown in Example 1, Step 7.

Example 3 – Download the same table and table B05001 for all counties in California

1. Let's assume that we have kept the DataFerrett session open while we looked at the data from Example 1 in our spreadsheet. We determine that this is the type of data we want, but we also discover that we need to add some data on the foreign born population. For that data we choose table B05001. We also decide that we want to get data for all the counties in California.
2. To select the data cells from B05001, all we need to do is go back to the Step 1 screen and repeat the actions shown in Example 1.
3. When we have finished these actions, we should see information for Foreign Born added to the list of the variables in the Step 2 screen (Figure 34):

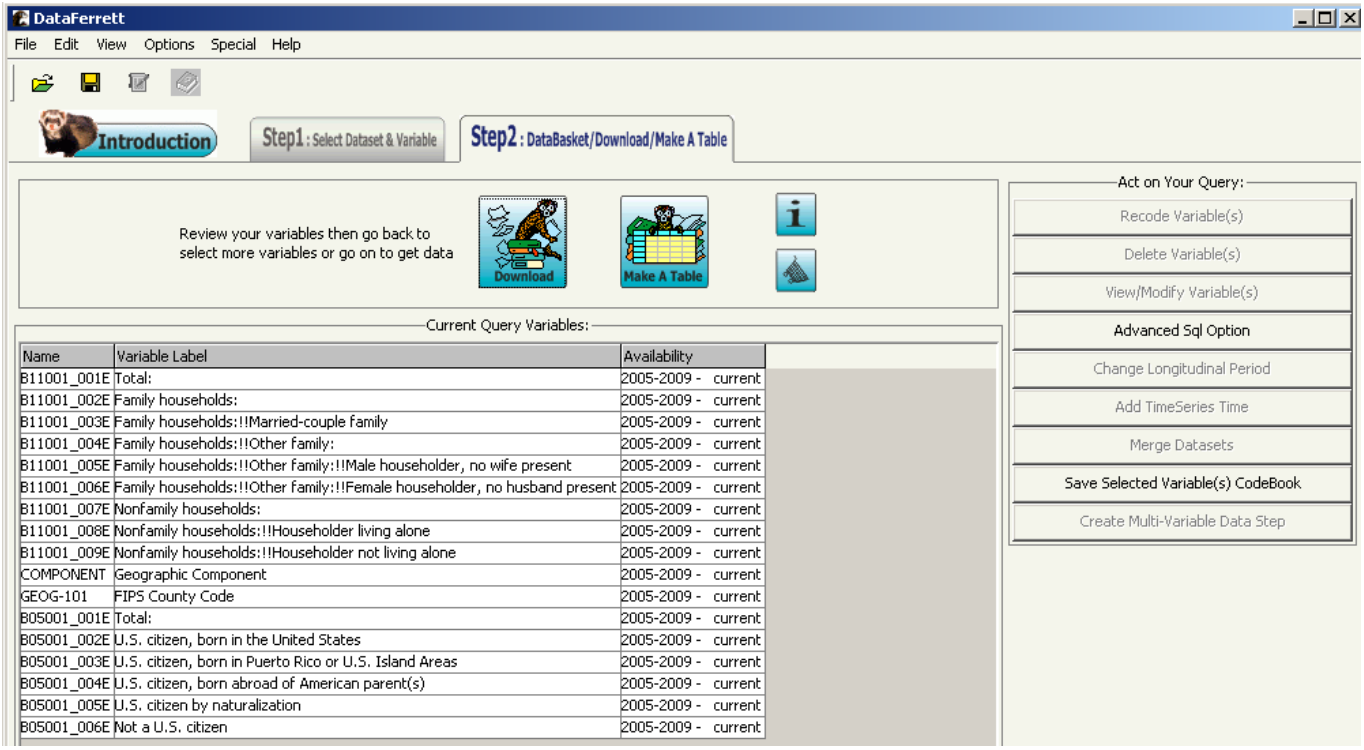

Figure 34: DataBasket With Added Variables for Foreign Born

4. Now we need to change our geography selection from Alameda County to all counties in California. To accomplish this, highlight the variables with “FIPS County Code” in the label and click on the “View/Modify Variables” button as indicated by the blue arrow below (Figure 35):

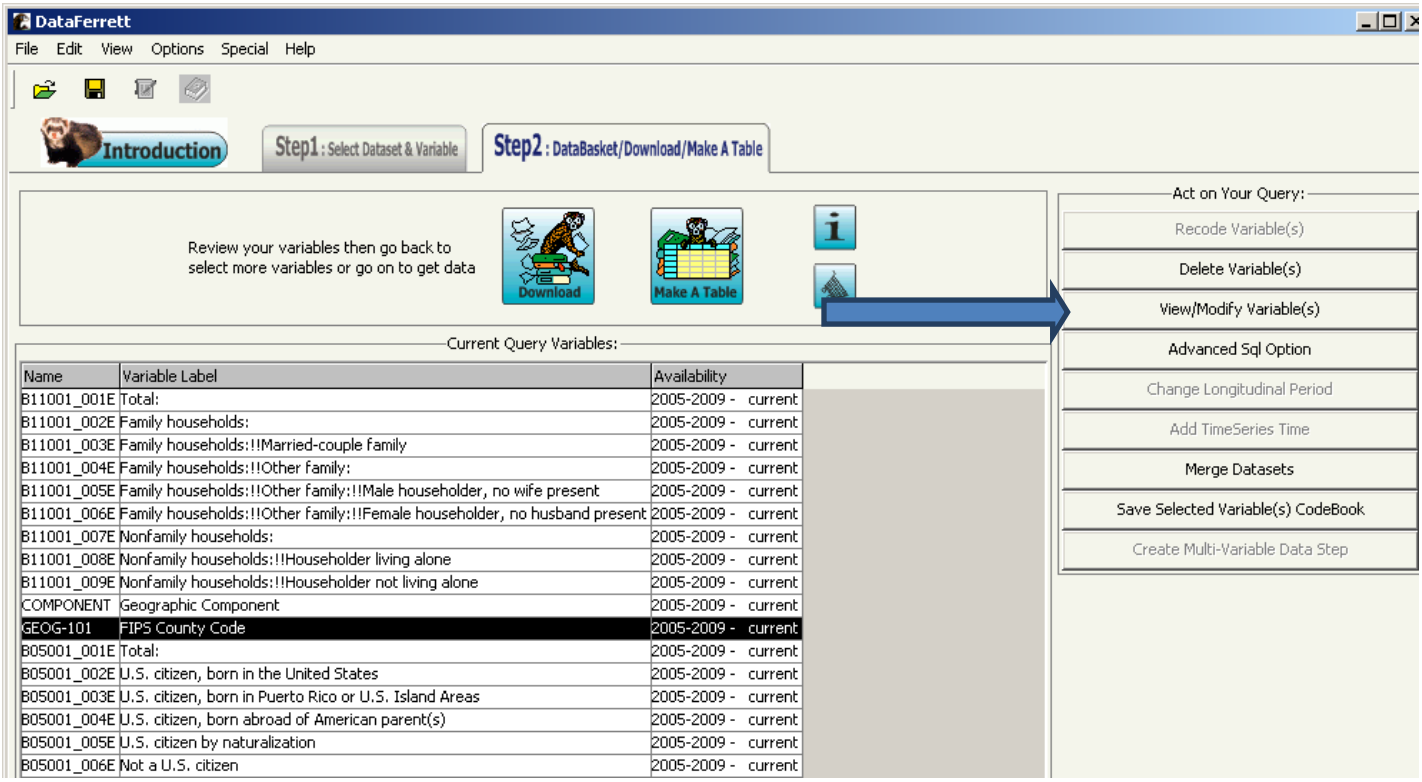

Figure 35: View/Modify Variables

5. Once you have clicked on the “View/Modify Variables” button for the county variable, you will be taken back to the “Ferrett Geography Codebook” window (Figure 36).

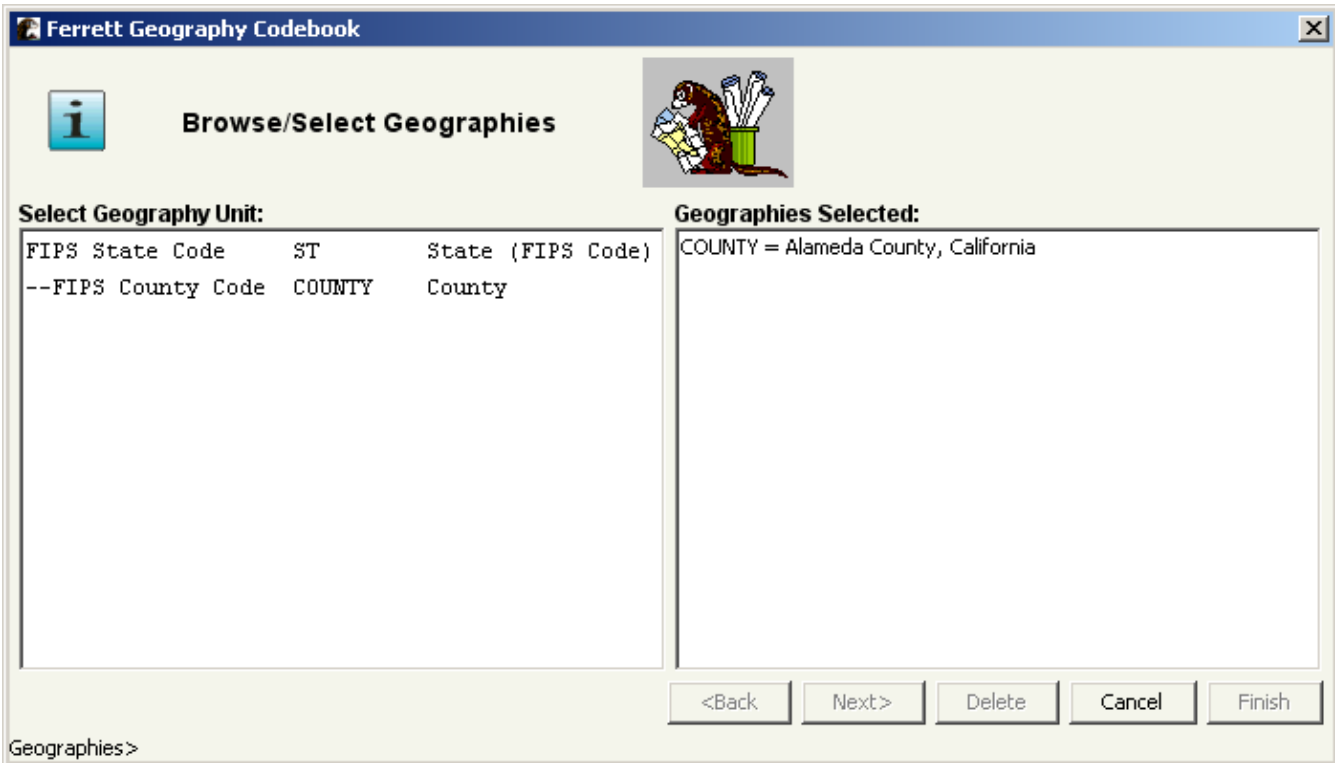

Figure 36: Ferrett Geography Codebook

6. Highlight Alameda County in the right-hand side of the window and click on the “Delete” button. You can now re-select counties as you did in Figure 15 of Example 1.
   a. You will see the screen shown previously but this time click on “Select All” at the top of the list of counties in California.
   b. Now you should see a listing of all the counties in California in the County Selected column on the right in the screen below (Figure 37). Click on the “Finish” button to complete the process.

Figure 37: Select County from List

c. At this point, you have made the changes to your DataBasket that you need; the dataset can now be downloaded your data as shown in Example 1.

Example 4 – Creating a thematic map

This example will demonstrate a few basic features of the "Make a Table" icon at the top of the screen in Step 2. We will use data cells we selected for Example 3 to create the thematic map showing the percent of foreign born who are not citizens in California. As a result, our starting point is very similar to Figure 34 in example 3. In Figure 38 below, under Step 2, instead of clicking on "Download", click on the "Make a Table" icon:

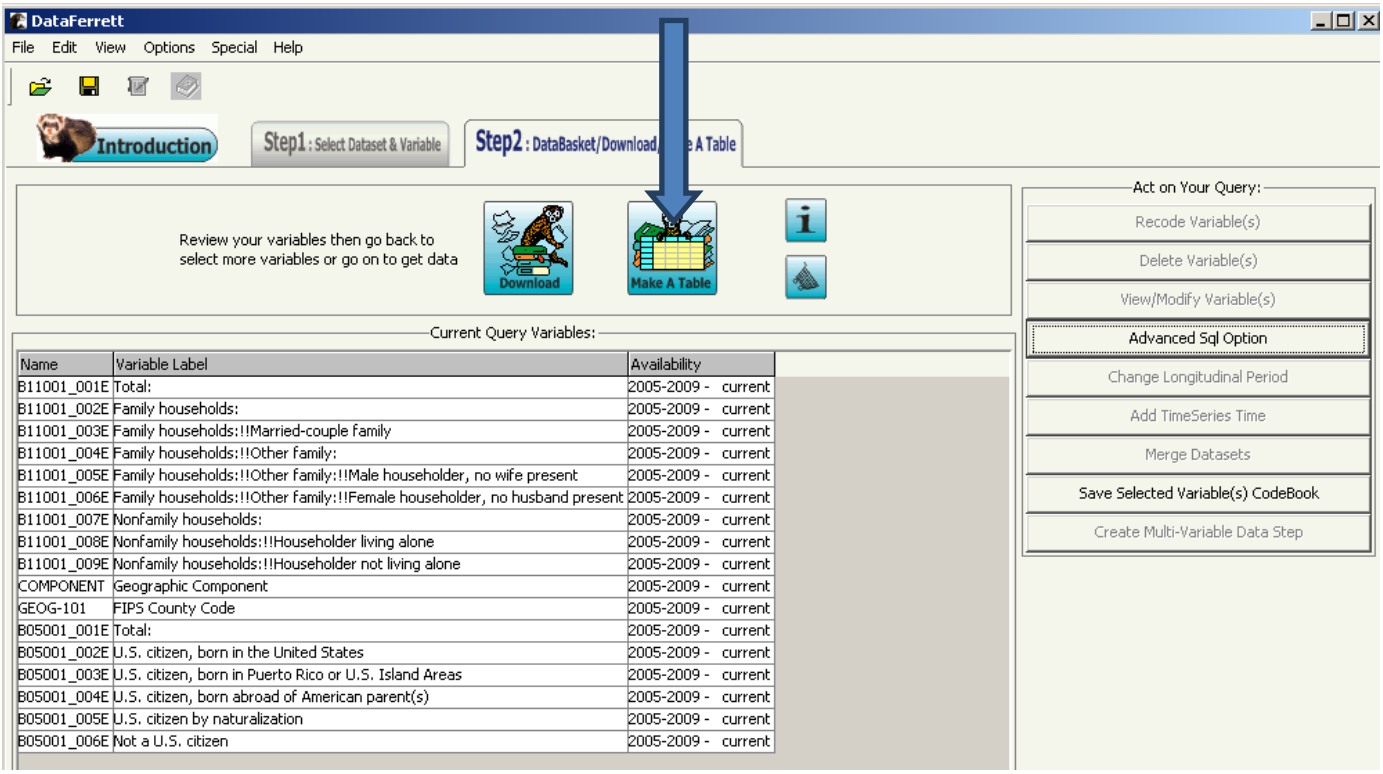

Figure 38: Make a Table Icon

1. After clicking on the “Make a Table” icon, the DataFerrett Tabulation window opens. To create a table, click on the “x” in the upper right hand corner of the warning screen that appears when you first open the table then select the variables that will populate the table.
2. The first variable will be FIPS County Code (GEOG-101) which should be placed in the first column.
3. Repeat the same actions with “Total” and “Not a U.S. citizen” to arrive at the following screen (Figure 39):

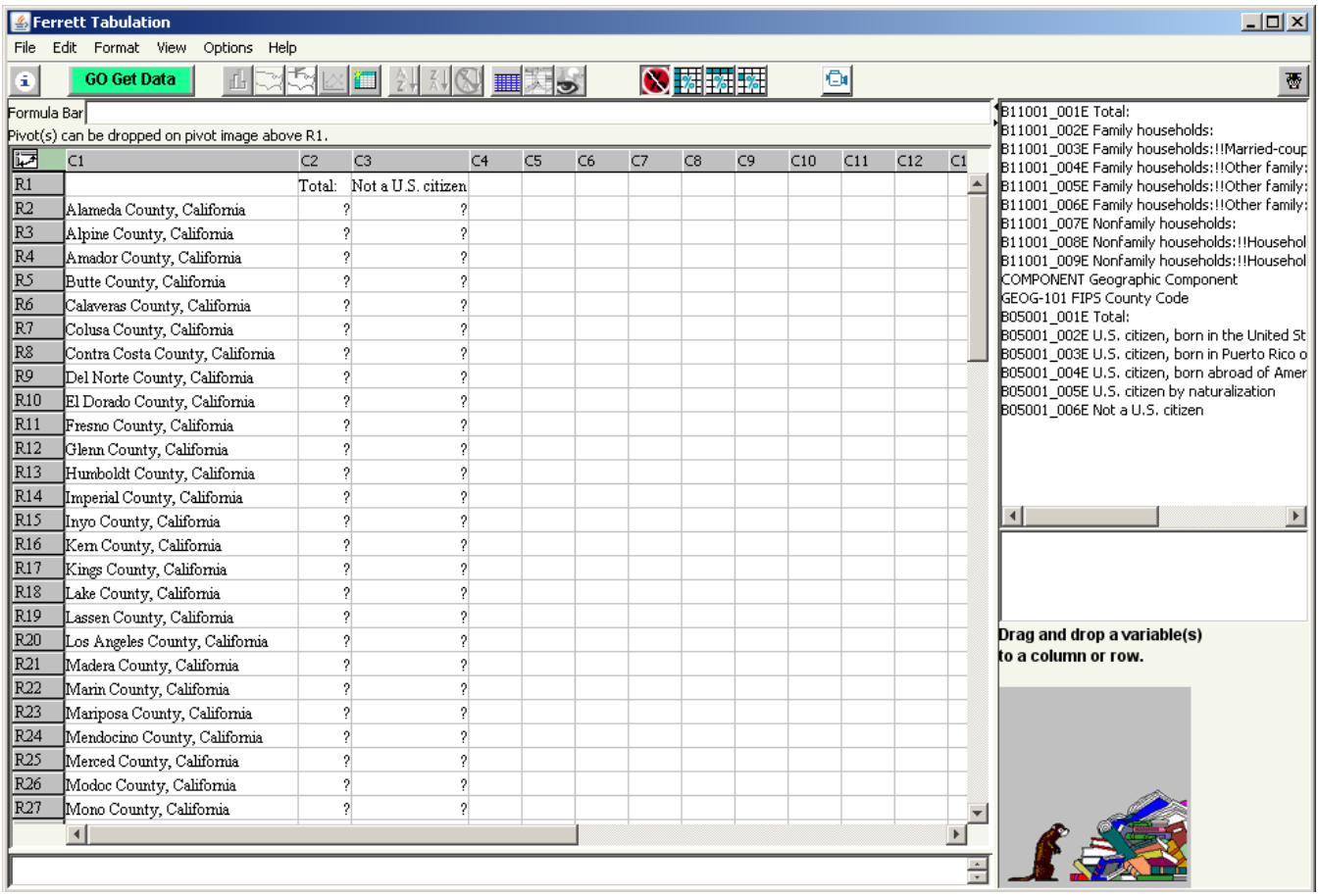

Figure 39: Make a Table Screen

4. Once the table is set up in the desired format, click on the “Go Get Data” button (pointed to by the arrow in Figure 40).

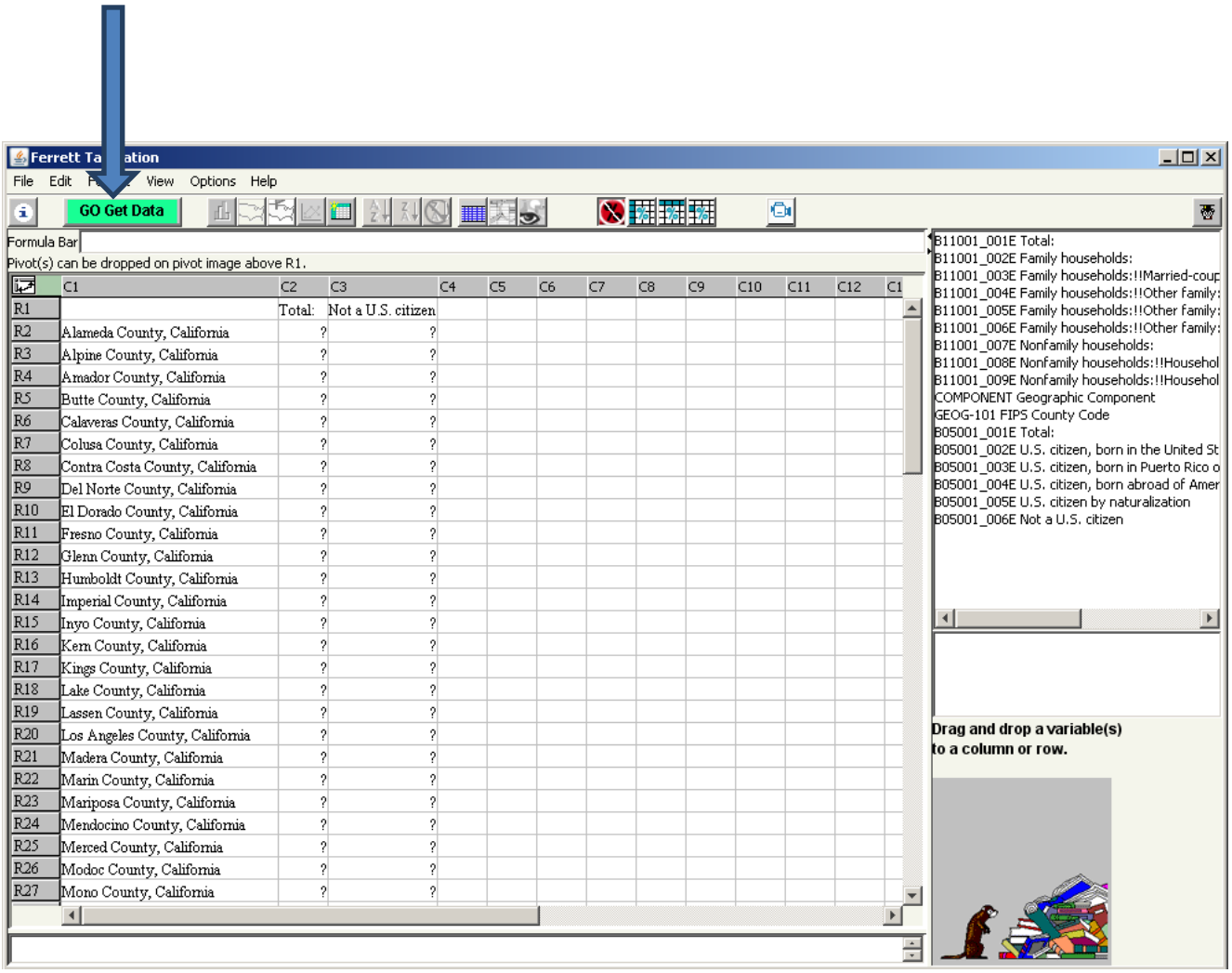

Figure 40: Select "GO Get Data"

5. The following screenshot (Figure 41) displays the tabulation window after the cells are populated with data. Next, convert the numbers in column 3 to a percent of the numbers in column 2 by clicking where the blue arrow is pointing in Figure 41.

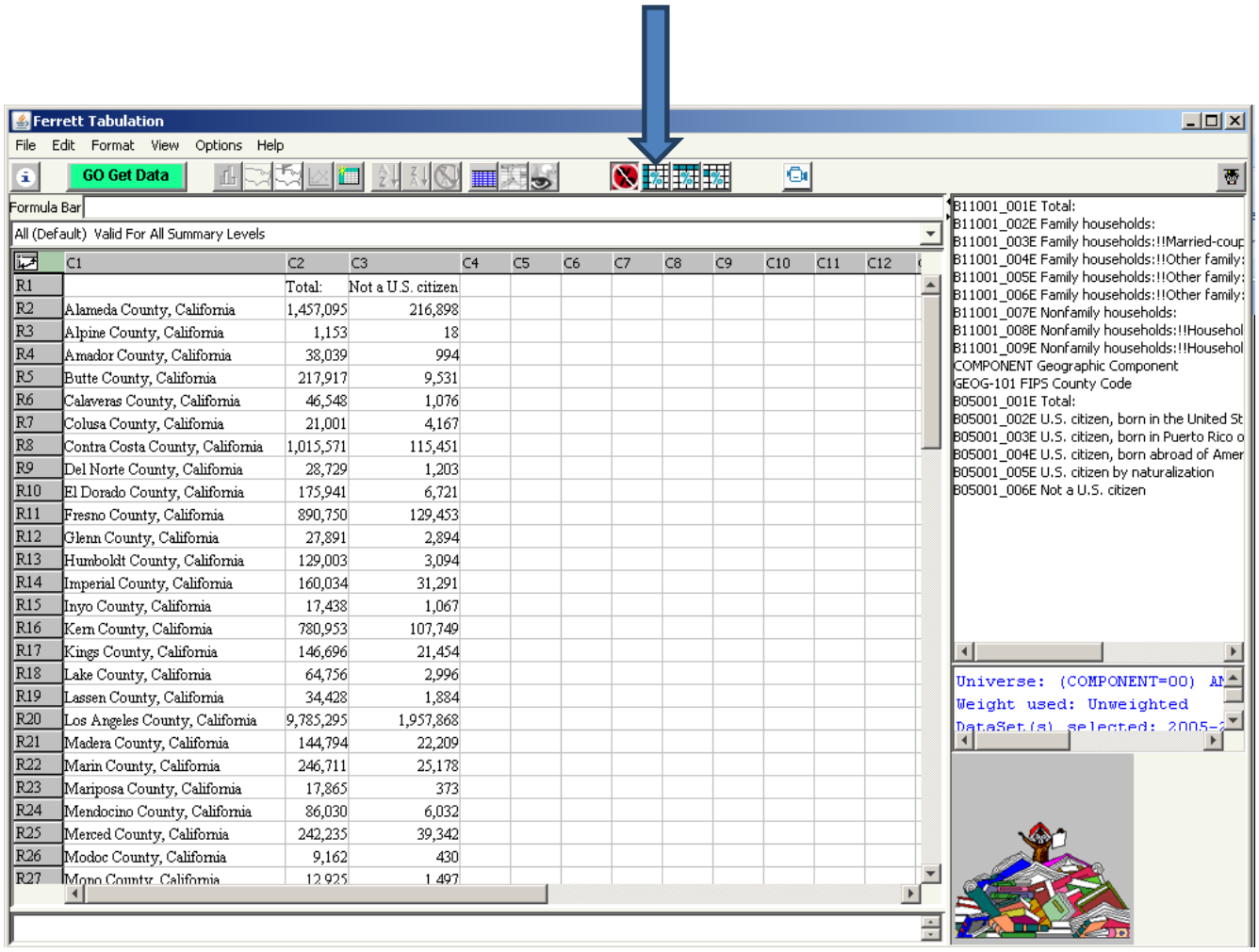

Figure 41: Convert Data to Percentage

6. Finally, highlight all the data cells in column C3. Do this by clicking on the cell inR2C3, holding down the SHIFT key, and scrolling to the last data cell in column C3 and clicking on it. The map icon at the top of the tabulation window (little yellow picture of the U.S.) will now be active (Figure 42). Click on the map icon and the map will appear as shown in Figure 43 below).

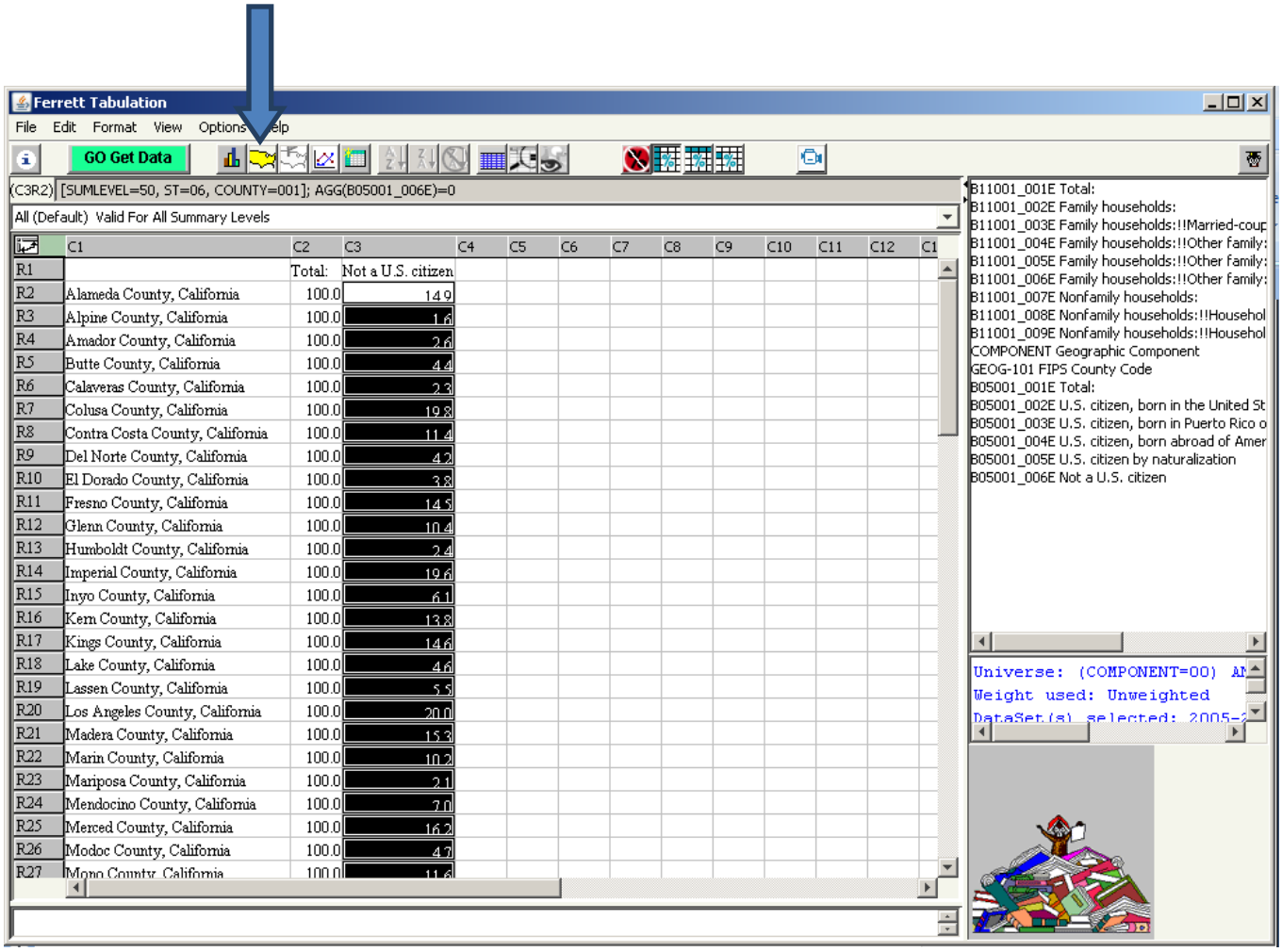

Figure 42: "Make a Map" Icon is Activated

7. The map is now displayed.

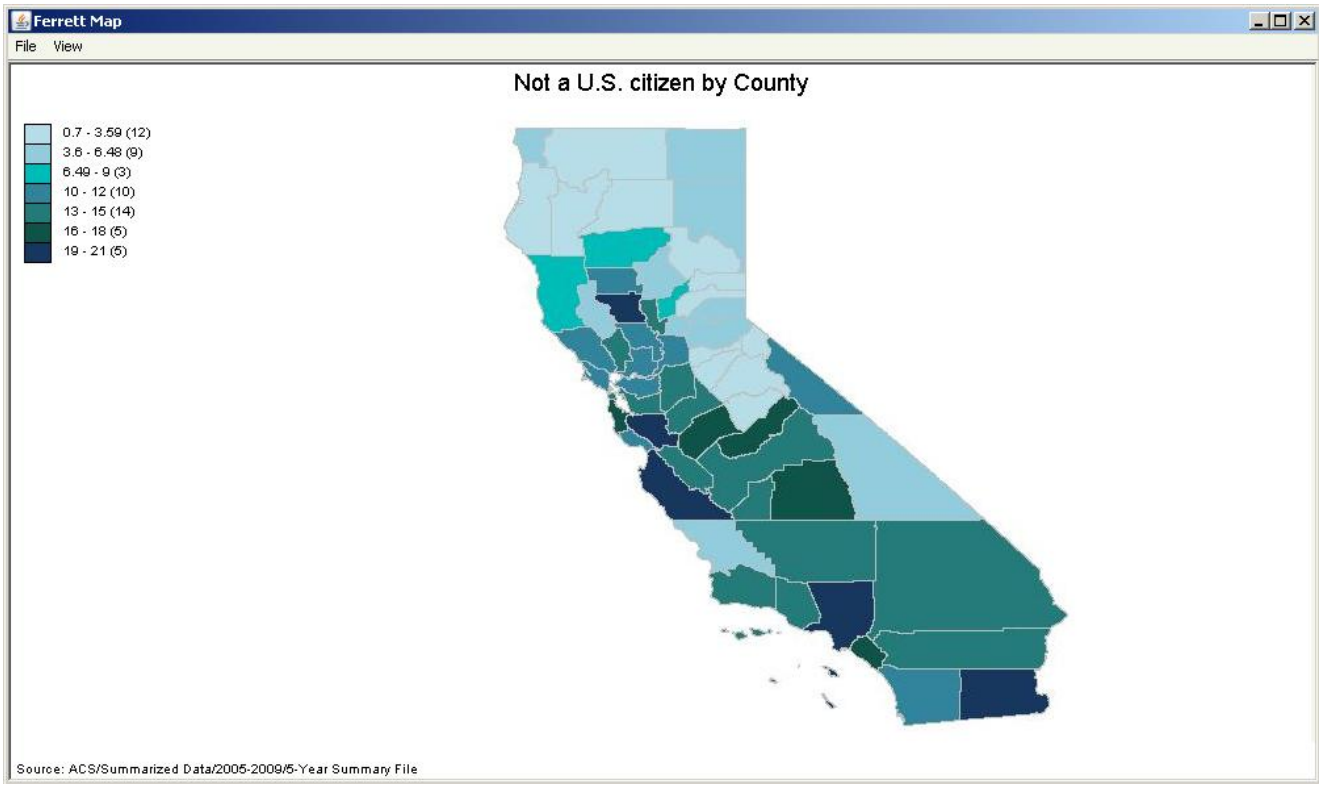

Figure 43: Completed Map for Not a U.S. Citizen by County

Two things to notes about this map:

a. In the map window, boundary layers can be added by clicking on the "View" menu tab and highlighting "Layers" at the bottom of the menu list. This displays the boundary choices that are available. In this example, you can add county boundaries to see the specific demarcation of each county.
b. Geography names are not available in the "Make a Map" option at this time.

New Approach to Displaying Margins of Error in Make-a-Table

Whether you are planning on downloading an extract or creating a table, the latest version of the ACS 5-year Summary File now allows the user to include Margins of Error (MOEs) as attributes of the estimates. In other words, the MOEs are no longer displayed as variables; the relationship between the estimates and the corresponding MOEs is maintained in DataFerrett.

Because DataFerrett treats the MOEs as attributes of estimates, the behavior of user-defined tables in the Make-a-Table window has changed.

There is no change in the way a user defines (lays out) a table. However, there are some changes to the Options menu for this window. Here is what the Options menu looks like after a basic table has been defined and populated with data:

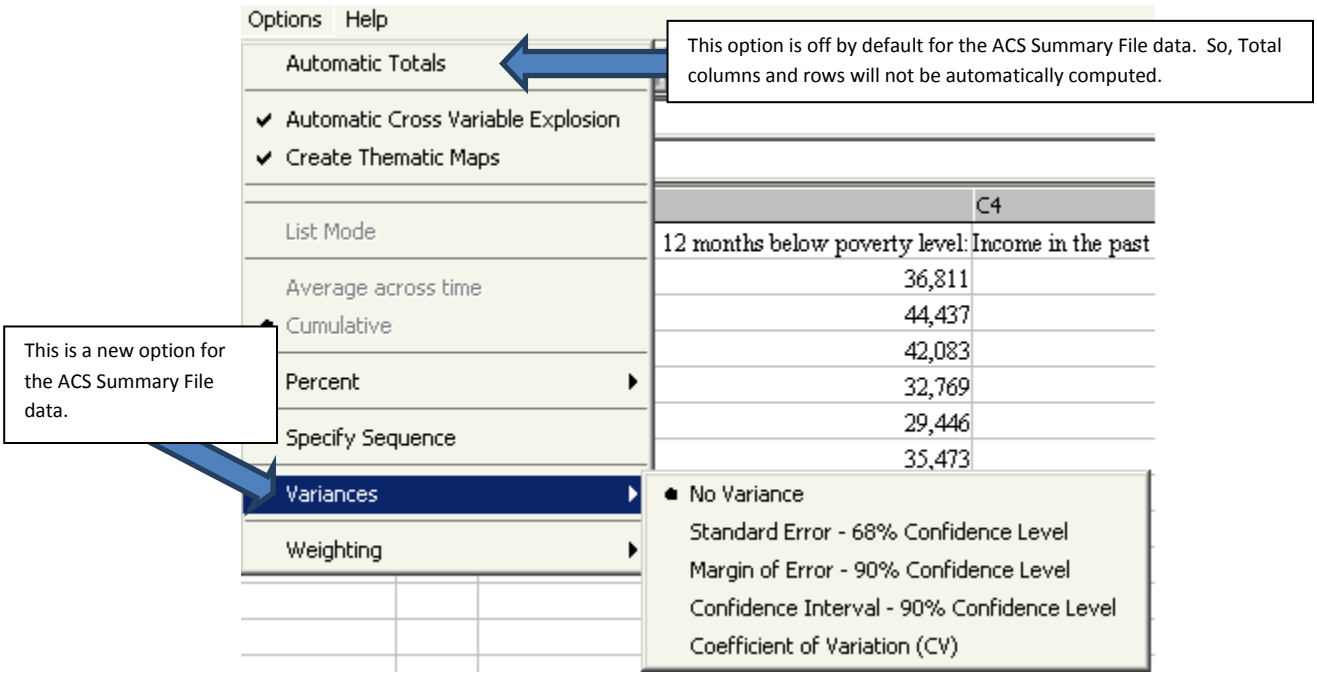

Figure 44: Options Menu in "Make a Table"

If the user turns on the "Margin of Error – 90% Confidence Level" Variances option, the table will be re-displayed with the 90% MOEs in parentheses after the estimates (in the same table cell) as shown below in Figure 45.

		C2	C3	C4	C
R1		Total:	Incomebelow poverty level:	Income at or above poverty level:	
R2	ct 01, Maryland	556,724 (2,767)	36,811 (1,694)	519,913 (2,979)	
R3	ct 02, Maryland	526,972 (3,985)	44,437 (2,062)	482,535 (4,057)	
R4	ct 03, Maryland	540,926 (3,796)	42,083 (2,189)	498,843 (3,813)	
R5	ct 04, Maryland	524,003 (2,566)	32,769 (1,836)	491,234 (2,935)	
R6	ct 05, Maryland	558,239 (2,722)	29,446 (1,666)	528,793 (2,824)	
R7	ct 06, Maryland	550,565 (2,208)	35,473 (1,531)	515,092 (2,421)	
R8	ct 07, Maryland	505,478 (3,120)	73,022 (2,841)	432,456 (3,635)	
R9	ct 08, Maryland	551,996 (2,708)	33,625 (1,921)	518,371 (3,258)	
R10					

Figure 45: Table with 90% MOEs

If, while the Variances option is on, the user turns on the "Automatic Totals" option and populates a table, this message will appear if the user does not include the "Total:" cell of a table. The message appears because DataFerrett actually calculates an approximation of the variance (MOE) under these circumstances.

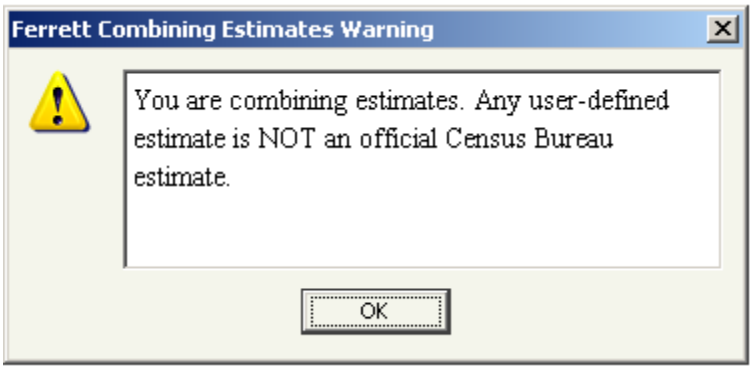

Figure 46: Ferrett Combining Estimates Warning

Other Features of the ACS Summary File in DataFerrett

Geography Components (Census defined geographic summary levels)

Geographic Summary Levels: There are 24 types of geography (aka "summary levels") currently available down to Block Group. Other summary levels will be added in later releases. Table 1 provides information on all geographic summary levels available in DataFerrett.

Table 1: Geographic Summary Levels

Type of Geographic Area - Code[1,4]	Type of Geographic Area - Description	Thematic Mapping available?[2,3]	Components (see Appendix F of the ACS Summary File Tech Doc for details)
10	United States	No	All
20	Regions	No	All
30	Divisions	No	All
40	States	Yes (within U.S.)	All
50	Counties	Yes (within state and U.S.)	No
60	County Subdivisions	Yes (within county and state)	No
140	Census Tracts	Yes (within county and state)	No
150	Census Block Groups	Yes (within tract and county)	No
160	Places	Yes (within state)	No
250	American Indian Area/Alaska Native Area/Hawaiian Home Land	Yes (within U.S.)	No
310	Metropolitan Statistical Area/Micropolitan Statistical Area	Yes (within U.S.)	No

320	State-Metropolitan Statistical Area/Micropolitan Statistical Area	Yes (within state and U.S.)	No
330	Combined Statistical Area	Yes (within U.S.)	No
340	State-Combined Statistical Area	Yes (within state and U.S.)	No
400	Urbanized Area	Yes	No
350	New England City and Town Area	Yes (within U.S.)	No
500	Congressional Districts - 111th	Yes (within state and U.S.)	No
510	Congressional Districts - 111th - County	Yes (within state)	No
610	State-State Legislative District (Upper Chamber)	Yes (within state)	No
620	State-State Legislative District (Lower Chamber)	Yes (within state)	No
795	Public Use Microdata Areas (PUMAs)	Yes (within state and U.S.)	No
950	Elementary School Districts	Yes (within state)	No
960	Secondary School Districts	Yes (within state)	No
970	Unified School Districts	Yes (within state)	No

NOTES:

1. The Census Bureau publishes 5-year estimates for many types of geographic areas (aka "geographic summary levels"). Many, but not all of these types of geographic areas are available in DataFerrett. Over time, the remaining types of geographic areas will be added to DataFerrett.
2. DataFerrett allows the user to create a map showing the distribution of values for any characteristic that can be represented in a column of a DataFerrett table for any type of geographic area which permits mapping (i.e., "Yes" in this column for the geographic area type).
3. Instructions for creating thematic maps in DataFerrett can be found in Chapter 14 of the User Guide, "Making a Map". (url: http://www.thedataweb.org/support/user/chapter14.html)

Figure 47 shows the list of all geographic **components** available in DataFerrett. Components are attributes of Census geography summary levels, for example, a user can limit the universe to the urban parts of a state by selecting the “01” component.

The “00” component is the default value, representing all of the geographic area(s) initially selected (Figure 47). You can add another component by clicking on the check box to the left of it and then clicking OK.

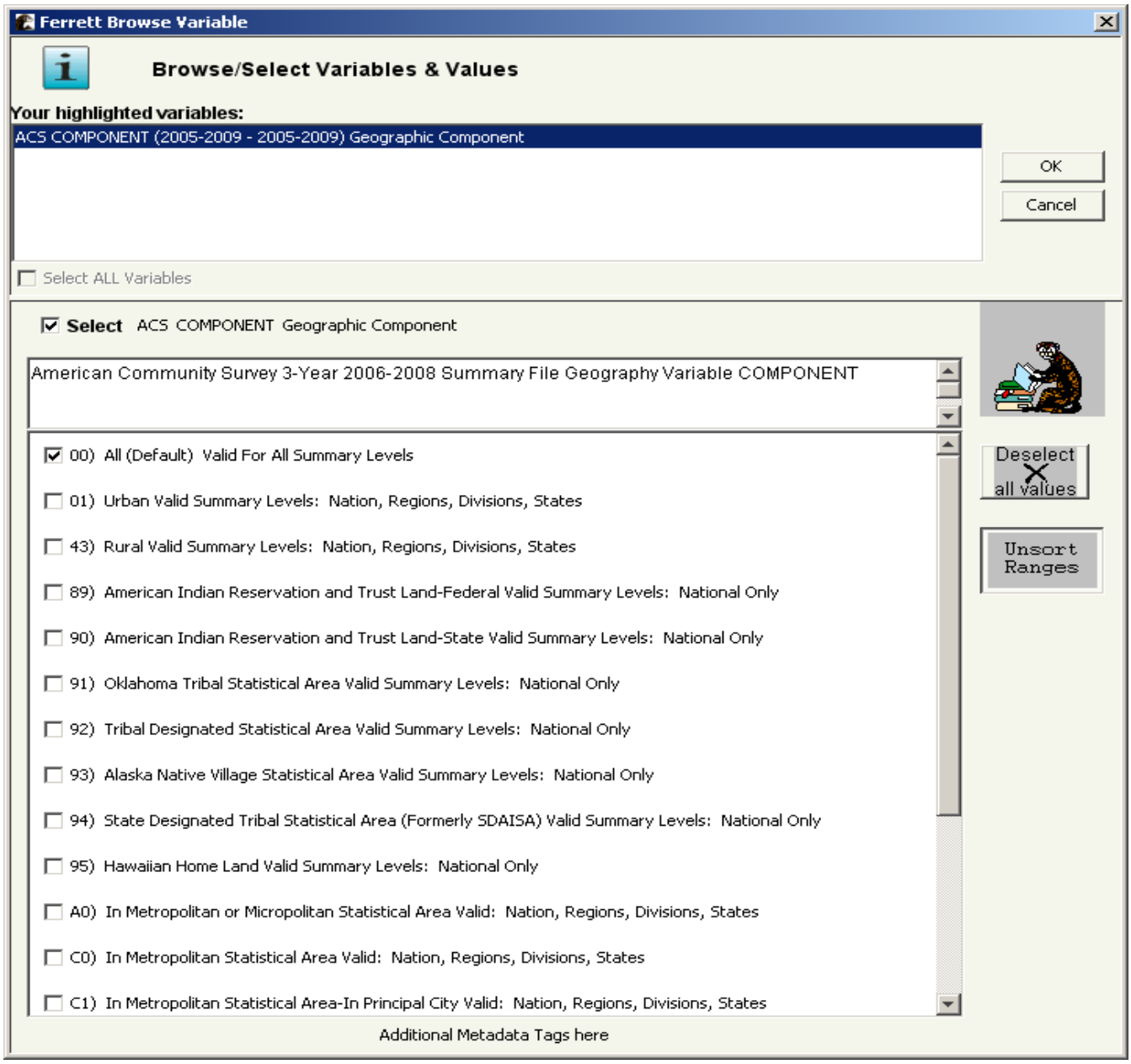

Figure 47: ACS COMPONENT Geographic Component

Tips for Using DataFerrett to Access the ACS 5-Year Summary File

Before you access DataFerrett:

1. Please review the documentation on the detailed tables in the 2006-2010 5-Year ACS Summary File in Appendix E of the technical documentation for the summary file located at http://www2.census.gov/acs2010_5yr/summaryfile/ACS_2006-2010_SF_Tech_Doc.pdf. This link points to a spreadsheet containing valuable information about the 940 detailed tables and the geographic areas for which each table is published. This is particularly important for users who want to create an extract containing data at the block group level, since only 338 of these tables exist at that level.

2. The "Derived Measure" tables (medians, ratios, aggregates, etc.) may have special values for certain cases in which no real estimate can be computed. These are the so-called "jam values" and are covered in Appendix C, section 3 of the technical documentation for the Summary File (see the link referenced in item 1 above).

3. All estimates are stored in DataFerrett as numeric values with these exceptions:
    - "null" is used when there is no estimate and the output is in comma-delimited or tab-delimited format.
    - Period (".") value is used for extracts in the SAS data format when there is no estimate.

When using DataFerrett to access the 5-Year ACS Summary File:

1. It is very important that you accept the security certificate that is displayed when accessing DataFerrett for the first time. Only then will all the features of DataFerrett related to saving files be available to you.
2. It is very important that you provide your correct email address to DataFerrett. All requests for data extracts from the ACS Summary File are run in a queued, non-interactive mode, and an email message is sent to the user when the extract is done and ready to be downloaded.
3. DataFerrett allows you to create very large extracts from the summary file data, but it is not designed to enable a user to download the entire summary file. Keep these restrictions in mind when specifying a data extract for downloading:
    a. The absolute limit of the number of variables that can be in your data basket when you request an extract is 535.

b. The absolute limit for the number of geographic areas that can be included in a single extract is 74,100. This number is slightly more than the total number of 2010 Census tracts in the U.S.
c. If you do select 74,100 geographic areas, you must keep the total number of variables in your extract at or below 200.
d. Using limitations b. and c., the largest number of cells allowed in a single extract is 200 X 74,100 or 14,820,000 cells.
e. If an extract exceeds 10 Megabytes in size but is below the other limits described above, you will be required to download the extract as a compressed file. If you check "PKZIP" as the type of compression, then you can use an application, such as WinZip, to uncompress the file once it is downloaded to your computer.

4. If you use the "**Make a Table**" feature of DataFerrett to look at all or some of the variables in your DataBasket, keep in mind that the default display is whole numbers for all numeric values in the Make-a-Table tabulation window. If a data column is defined to have accuracy at a decimal level, such as tenths, you must highlight the column, click on the Format drop-down menu, click on "Decimal", and choose the accuracy level needed. The display will be updated with the correct accuracy once you have done this.
5. Example 4 in this document contains instructions on how to use the "**Make a Table"** feature of DataFerrett. If you are not familiar with this feature, it is highly recommended that you read these instructions before trying create a table in DataFerrett.
6. All extracts of the summary file data include the GEOID variable as the main geographic identifier for each record. For those users not familiar with this variable, see section 2.6 of http://www2.census.gov/acs2010_5yr/summaryfile/ACS_2006-2010_SF_Tech_Doc.pdf.
7. It can happen that a user action in DataFerrett will force the display of a message in html format using the user's web browser. To return to the DataFerrett application, the user can find it on the task bar and click on it. In most cases, the icon will look like this: